Final Technical Report Template

Final Technical Report

Costs of Edaphic Stress to the Australian Grains Industry

Project code:	UQ00081
Prepared by:	Thomas Orton t.orton@uq.edu.au The University of Queensland Yash Dang y.dang@uq.edu.au
Date submitted to GRDC:	29/09/2017

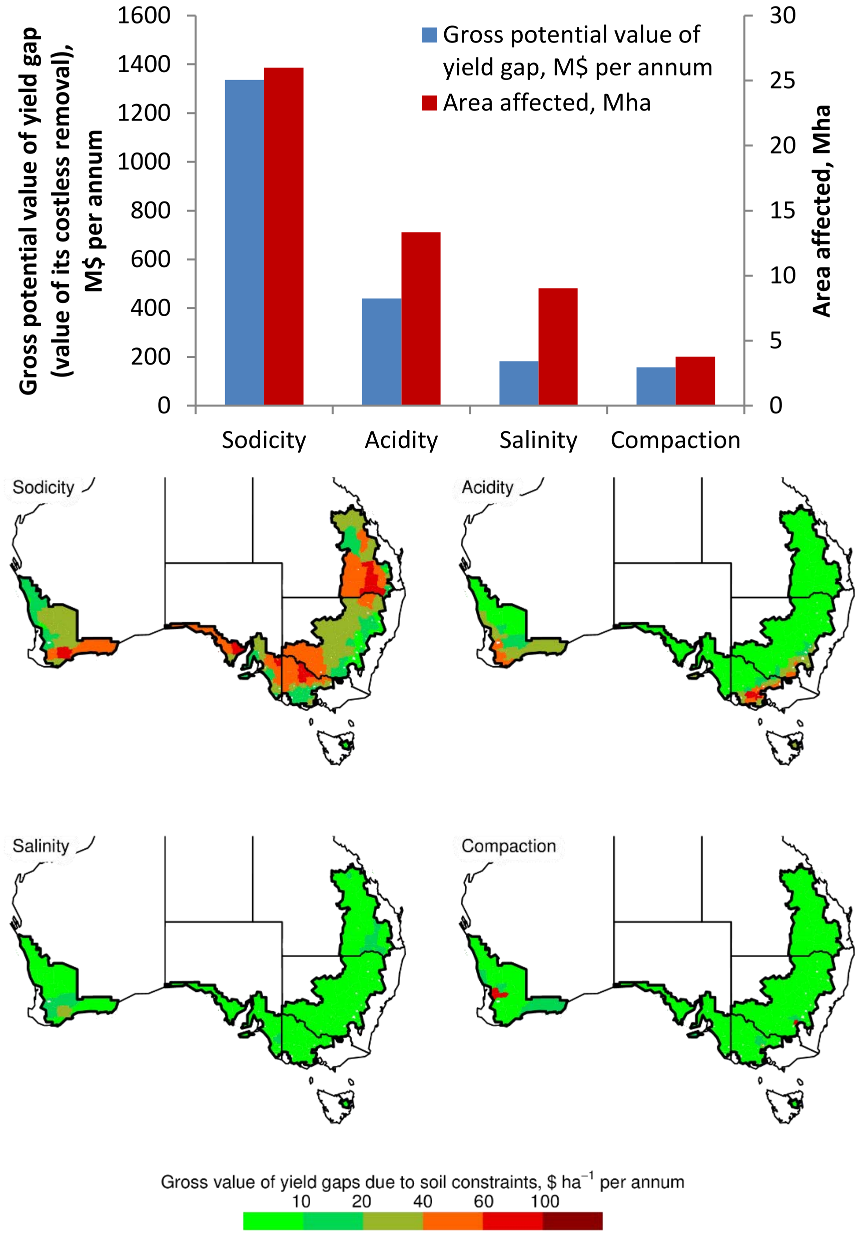
REPORT SENSITIVITY

Does the report have any of the following sensitivities?

Intended for journal publication	YES
Results are incomplete	NO
Commercial/IP concerns	NO
Embargo date	NO

Abstract

Soil sodicity, acidity, salinity and compaction are important constraints to grain production on many cropping soils in Australia. Each soil constraint has different causes and management options; this presents an economic opportunity, the extent of which varies spatially given the highly variable nature of soils and management practices. Despite their known significance, there is considerable variation in the existing information on the costs of each soil constraint to Australian agriculture. Such accurate information is vital to GRDC to guide investment decisions on remediation and minimise productivity losses. This project presents and applies a framework for developing national maps quantifying forfeited grain yields due to specific soil constraints (sodicity, acidity, salinity and compaction) at a broad spatial scale and for assessing the economic benefit of ameliorating or managing these constraints. The approach brings together data from diverse sources, with careful consideration of their different spatial scales, to allow the formulation of models that can predict lost yield due to soil constraints. Of the four constraints considered, results indicate that sodicity gave the largest magnitude of yield gaps across most regions. Yield gaps due to acidity were more concentrated spatially, in the high-rainfall regions of Western Australia, Victoria and New South Wales.



Executive Summary

Soil sodicity, acidity, salinity and compaction are important edaphic (soil-related) constraints to grain production on many cropping soils in Australia. Each soil constraint has different inherent or management-influenced causes and their management presents an economic opportunity, the extent of which varies spatially given the highly variable nature of soils and management practices. Despite their known significance, the available information on the extent and impact of these constraints on the grains industry is mostly based on extrapolations from soil surveys and expert opinion from individual regions. Hence, there is considerable variation in the existing information on the costs of each of the soil constraints to Australian agriculture. Therefore, the Australian grains industry needs accurate information on the costs of these constraints at a national scale to guide investment decisions on remediation to minimise productivity losses.

Estimating the severity of soil constraints, and their impact on management and plant productivity, is a very complex issue. Several surface and subsoil properties interact with each other to determine the local environment for root growth at a given time. Rarely do the various soil constraints occur independently. Moreover, the variable distribution of soil constraints, both spatially within a paddock, across the landscape and down the soil profile, and the complex interactions that exist among the various physical and chemical soil properties, make it difficult to determine which constraint is the major limitation to crop production at any given location. As management practices and plant growth vary significantly under different weather patterns, the influence of these soil constraints on production also varies across seasons. In most cases, the impact of soil constraints has been evaluated separately in both surface and subsoils and often there is a debate about the relative importance of different soil constraints to agricultural productivity. Identification of the most-limiting constraint and its interaction with other factors is a first step in planning for soil management and the selection of traits for the breeding of adapted crop cultivars. The objectives of this project were to:

(i) develop national maps that quantify forfeited grain yields due to specific soil constraints of sodicity, acidity, salinity and compaction at the level of Statistical Area Level 2 (SA2);
(ii) develop a framework for assessing the economic-benefit of ameliorating and/or managing specific soil constraints.

These objectives have been accomplished through a multi-stage approach, which has brought together data from a number of diverse sources: grain yield data at the SLA (Statistical Local Area) level from the Australian Bureau of Statistics, remote sensing data on 30-m and 250-m pixels from the Landsat and MODIS satellites, climate data on a 5-km grid across Australia from the Scientific Information for Land Owners (SILO) database hosted by Queensland Government, and soil data from soil profiles across Australia's cropping land, predominantly from the National Soil Site Collation. Information from these diverse sources has been combined, with careful consideration of their different spatial scales, in such a way as to allow the formulation of models that can predict the lost yield due to soil constraints. Subsequently, geostatistical approaches have been applied to convert the spatial scale of the yield gap estimates to the desired reporting level of SA2.

Key results for the four main soil constraints we considered in this project are:

Soil sodicity Sodic soils are characterised by an excess of sodium on exchange sites (> 6%), and lead to soil dispersion, decline in soil structure, low water infiltration and high susceptibility to erosion. Soil sodicity is estimated to affect 26 million ha across all grain-cropping regions of Australia. Its costless removal is estimated to be worth $1.34 billion per annum, and the predicted potential benefits of ameliorating soil sodicity through gypsum application (based on indicative treatment costs) are $1.15 billion per annum.

Soil acidity Acidic soils are characterised by soil pH < 6.0 (as measured in 1:5 0.01 M calcium chloride solution) in the 0–10 cm soil depth and < 4.8 in subsoils (below 10 cm), and create chemically undesirable conditions for many crop species. Soil acidity is estimated to affect 13 million ha mainly in Western Australia, Victoria and southern New South Wales. Its costless removal is estimated to be worth $440 million per annum, and the predicted potential benefits of ameliorating soil acidity through lime application (based on indicative costs) are $380 million per annum.

Soil salinity Saline soils are characterised by an electrical conductivity of > 0.3 dS m^{-1} in the 0–10 cm soil depth or > 0.7 dS m^{-1} in subsoils. They are also characterised by a chloride concentration of > 300 mg kg^{-1} in the 0–10 cm soil depth or > 600 mg kg^{-1} in subsoil, which provides toxic conditions for many crop species. Soil salinity is estimated affect 9 million ha of currently cropped land, mainly in Western Australia and southern Queensland, and its costless removal is estimated to be worth $180 million per annum. However, it should be noted that our estimates are based on currently cropped land, and do not include land which has entirely lost its potential for cropping due to salinity, which can be the case in particular in Western Australia.

Soil compaction Soil compaction can have a detrimental effect on root growth when bulk density exceeds 1.6 g cm^{-3}. Soil compaction is estimated to affect 4 million ha and its costless removal estimated to be worth $160 million per annum.

Validation of areas affected and of yield gaps due to soil constraints has been carried out using data from four sources, including the GRDC's National Paddock Survey project. Results showed modest agreement in terms of areas affected, though differences between calibration and validation datasets and between the definitions of 'areas affected' were probable causes of this. Validation of yield gaps due to soil constraints also showed some agreement between predicted and observed values, but a limitation in validation data—where yields for paired samples, one affected and one unaffected by the soil constraints, were available—meant that robust conclusions could not be drawn. A sensitivity analysis was carried out and results suggested that sodicity, even when defined using sodic-tolerant critical values, resulted in the largest yield gaps. Also, an alternative to the empirical modelling approach to estimate yield gaps due to soil constraints—based on the mechanistic agricultural systems simulation modelling software, APSIM—has been demonstrated.

A few caveats should be noted for the interpretation of results. We have presented results as maps of soil constraints at SA2 level, which should be interpreted with care. The statistical areas themselves are of vastly different sizes, and furthermore vary considerably in the intensity of cropping land. Therefore data presented in such maps may create biases in interpretation. To overcome this problem, data relating to maps are also tabulated and made available in an interactive Excel spreadsheet form, whereby one can select any particular SA2 of interest and explore its full set of data. This should provide a valuable resource for GRDC to draw robust insights from this study. A further implication of reporting at SA2 level is that variability of the effects of soil constraints within SA2s becomes averaged out; in some cases this could mask information, for instance where an SA2 is characterized by highly acidic upland sands and highly alkaline valley floors as can be the case in particular in Western Australia. It should also be noted that the methodology was designed to provide a nationally consistent picture of costs of edaphic stress by making the best use of data available and summarizing results at large spatial scales (SA2). There are a number of steps in the process, all of which carry some level of uncertainty and are intended to provide information at broad rather than fine spatial scale.

Of the four soil constraints considered in this work—sodicity, acidity, salinity and compaction—sodicity gave the largest magnitude of yield gaps across most regions of Australia, with an average yield gap of more than twice that of acidity. Yield gaps due to acidity were more concentrated spatially, in the high-rainfall regions of Western Australia, Victoria and New South Wales.

GRDC GRAINS RESEARCH & DEVELOPMENT CORPORATION

Contents

Abstract 3

Executive Summary 4

Contents 6

Background 7

Project objectives 7

Methodology 10

Results 30

Discussion of Results 41

Conclusion 44

Implications 45

Recommendations 46

The Project Team 47

Appendix A. Cubist modelling details 47

Appendix B. Cubist model rules and parameters 47

Appendix C. Sensitivity analysis results 47

Appendix D. APSIM modelling details 47

Glossary and Acronyms 60

References 63

Background

The 2012–2017 Strategic Research and Development Plan of the Grains Research and Development Corporation (GRDC) focused on protecting and enhancing the farms' resource base, in order to maintain or enhance production. GRDC has identified that, in spite of continued varietal improvement and other crop management practices, soil constraints are becoming a more obvious and significant limitation to yields and profits. The yield of many crops in Australia remains well below the average yield obtained in many western countries. For example, the average yield of wheat in Australia is ~1.7 t ha^{-1} as compared to the world average of 2.7 t ha^{-1}. In spite of many technological innovations and varietal improvements in crop production there is evidence that the rate of crop productivity gain has declined globally during 1990–2007 compared with 1961–1990 and this is more so in Australia (Kokic et al., 2006). To meet global food security there is an urgent need to increase food production either through vertical increase in yield per unit of land or through horizontal increase in the area of land deployed for agricultural production. The latter is not possible in much of Australia.

The National Land and Water Resources Audit (NLWRA 2001) found that the health of our soils that underpin agricultural productivity is deteriorating across Australia. The area of productive land in Australia is diminishing due to urbanisation and various soil degradation problems. It is thought that approximately 75% of Australian soils have single or multiple soil constraints in both surface and subsoil and this area is increasing over time (Bot *et al.* 2000). Limitations to agricultural productivity imposed by soil-related stresses are often severe and need remediation to improve grain yield—i.e. closing the gap between actual and potentially attainable yields—and thereby meet global food security requirements as well as maintain the profitability of Australian grain farms.

Soil sodicity, acidity, salinity and compaction are important edaphic (soil-related) constraints to grain production on many grains cropping soils in Australia. Each soil constraint has different causes and treatments. However, the management of each soil constraint presents a varying economic opportunity, the extent of which is not well known. Despite their significance, the available information on the extent and impact of these constraints on the grains industry is mostly based on extrapolations from soil surveys and expert opinion from individual regions; although these sources can provide valuable information for management advice, they are difficult to collate into a consistent national overview. There is considerable variation in the existing information on the costs of each of the soil constraints to Australian agriculture (Dang and Moody, 2016). Therefore, the Australian grains industry needs accurate and standardized nationwide information on the costs of these constraints at a national scale to guide investment decisions on remediation to minimise productivity losses and to set priorities for the selection of traits for the breeding of adapted crop cultivars.

Estimating the severity of soil constraints, and their impact on plant productivity and management is a very complex issue. Several surface and subsoil properties interact with each other to determine the local environment for root growth at a given time. Rarely do the various soil constraints occur independently. Moreover, the variable distribution of soil constraints—spatially within a paddock, across the landscape and within the depth of the soil profile—and the complex interactions that exist among the various physical and chemical soil properties, make it difficult to determine which constraint is the major limitation to crop production at any given location (Dang et al., 2010). As management practices and plant growth vary significantly under different weather patterns, the influence of these soil constraints on production also varies across seasons. In most cases, the impact of soil constraints has been evaluated separately for the surface and subsoils and often there is a debate about the relative importance of different soil constraints to agricultural productivity.

Previous work in the wheat belt of Australia (Hochmann et al., 2012) has sought to quantify the yield gap, Y_g, as the difference between yields that are currently achieved by farmers, Y_a, and those potentially attainable (by an adapted crop variety without growth limitations from nutrients, pests or diseases) under rainfed conditions, Y_w (the *water-limited* yield):

$$Y_g = Y_w - Y_a \quad (1)$$

The resulting yield gap was attributable to genetic and management factors. In the current study we are concerned with individual contributions to lost yield from specific soil constraints. Therefore, our approach and terminology differ from that of Hochmann et al. (2012). Nevertheless, the results obtained from the current study should be seen to complement those of previous studies, including Hochmann et al. (2012).

Project objectives

Given the variation in the existing information on the costs of each of the soil constraints to Australian agriculture, the objectives of this project were to:

(i) develop national maps that quantify forfeited grain yields due to specific soil constraints of sodicity, acidity, salinity and compaction at the level of Statistical Area Level 2 (SA2);

(ii) develop a framework for assessing the economic-benefit of ameliorating and/or managing specific soil constraints.

We reiterate that the target 'spatial support' (i.e. the area over which variables are represented as an average) for the information produced is the SA2 level of the Australian Statistical Geography Standard, ASGS (ABS, 2017). This work does not aim to provide individual farmers with advice on the soil constraints that most affect their own yields. Rather, the aims are to summarize by region what the most likely important constraints are and to provide some indication of the likely general benefits of remediation, using existing knowledge. Within an SA2 there will be considerable variation. Therefore, the scope of this work will be to provide GRDC with information about the relative significance of likely benefits of remediation at the SA2 level. More locally focused methods of investigation would be required to elucidate further the likely benefits to a land owner at a particular location under specific management practices and environmental conditions. For instance, should the current study identify an SA2 as having a large yield gap due to sodicity, further investigation of datasets, expert opinion and other background information relevant to sodic soils in that SA2 (e.g. digital terrain maps, local soil maps, other local studies and farmer knowledge) should be able to shed light on reasons for the large yield gap.

Methodology

Method overview and data

As noted, our terminology will differ slightly from that used previously for yield gap work in Australia (Hochmann et al., 2012). We will define the yield gap due to soil constraint '*c*' as:

$$Y_{gc} = Y_{oc} - Y_{ac} \tag{2}$$

Here, the yields that we will refer to as 'actual yields', Y_{ac}, are those that are predicted by a model (that represents the effects of soil constraint *c* on yield) with the actual soil and climate data as inputs, while the 'constraint-optimized yields', Y_{oc} are those predicted by the same model with the same climatic inputs but with soil constraint *c* set to some defined optimum value. Our task then was to formulate models that could represent the effects of soil constraints on yield across the variety of environmental conditions encountered in the wheat-growing regions of Australia.

We adopted an empirical modelling approach to represent the effects of soil constraints on yield. The procedure involved the following steps:

(i) Define active-cropping areas for each winter-wheat growing season (May–November, 1999–2012)
(ii) Disaggregate yield over soil data locations for each growing season
(iii) Harmonize soil data for each profile at three consistent depths
(iv) Define optimum ranges and the type of yield decline curves for each soil constraint
(v) Pre-process climate data
(vi) Fit model: yield = *f*(soil constraint *c*, other soil, climate)
(vii) Calculate at the soil data locations:
  a. Y_{ac} by applying the model with actual soil data
  b. Y_{oc} by applying the model with soil constraint *c* optimized
  c. Y_{gc} through Equation 2
(viii) Interpolate Y_{gc} to 1-km grid over entire cropping area
(ix) Aggregate predictions to SA2 level for reporting and economic analysis
(x) Undertake economic analysis at SA2 level

Our objectives have been accomplished through the above multi-stage approach that has brought together data from a number of diverse sources, as summarized in Table 1: wheat yield data at the Statistical Local Area (SLA) level from the Australian Bureau of Statistics (ABS)[1], remote sensing data on 30-m and 250-m pixels from the Landsat and MODIS (MOD13Q1) satellites, climate data on a 5-km grid across Australia from the Scientific Information for Land Owners (SILO; Jeffrey et al., 2001) database hosted by Queensland Government, and soil data from soil profiles across Australia's cropping land, predominantly from the National Soil Site Collation (NSSC). Information from these diverse sources has been combined, with careful consideration of the different spatial scales, in such a way as to allow the formulation of models that can predict the lost yield due to soil constraints. The economic analysis is the final step in the process; it is applied at the SA2 level and utilizes as its input data the output from steps (i)–(ix) that estimate the magnitudes of yield gaps due to soil constraints. In the following section we provide details about what was involved in each of the above steps.

[1] The ABS has recently replaced its Australian Standard Geographical Classification (ASGC, of which the SLA is a defined area) by the Australian Statistical Geography Standard (ASGS, of which the SA2 is a defined area). At the time of starting the project, ABS and CSIRO were using SLAs as the unit by which to summarize yield data, and although they have recently changed to SA2s, the data available for undertaking this work were at SLA level.

Table 1 A summary of data sources used to derive estimates of yield gaps due to soil constraints.

Variables	Spatial/temporal data support	Grid size/number of data	Source
Land use; dryland cropping, irrigated cropping, other	Point (each pixel's class observation assumed to be representative class for the entire 100-m × 100-m pixel)	On 100-m × 100-m grid	ACLUMP; http://www.agriculture.gov.au/abares/aclump
EVI	30-m × 30-m (Landsat) and 250-m × 250-m (MODIS) pixels (value of pixel assumed to be representative of entire pixel); point-in-time measurements	On 30-m × 30-m (Landsat) and 250-m × 250-m (MODIS) grids; measurements approximately every 16 days from both Landsat and MODIS, those from MODIS representing the maximum EVI in a 16-day window from overpasses every 1–2 days	Landsat: USGS; https://landsat.usgs.gov/ MODIS: USGS; https://lpdaac.usgs.gov/dataset_discovery/modis
Yield	Averages over the active cropping areas of SLAs, which vary in area of potential cropping from < 500 ha to 1.3 million ha	254 SLAs with wheat yield data across Australia, each with one average value for each winter-wheat growing season from 1999 to 2012	ABS and CSIRO; http://www.yieldgapaustralia.com.au/wordpress/ (Note that the above source has recently updated the spatial support of yield data from SLA to SA2; we work with SLA-level data that were available at the time of commencing the project)
Climate; VPD	Point (each pixel's value assumed to be representative of entire 5-km × 5-km pixel); point-in-time measurements	On 5-km × 5-km grid; daily measurements over the period September–October for each growing season from 1999 to 2012	SILO (gridded dataset); https://www.longpaddock.qld.gov.au/silo/index.html
Soil; pH, ESP, EC, Cl, BD, sand, clay	Point, each from a single soil profile. Each soil profile consists of a number of depth intervals (an average of 4 depth intervals per profile), each measurement representing the average of the soil property over that interval	Data from 30549 depth intervals within 7015 soil profiles; varying numbers of data for between soil properties	Originally from State and Territory agencies, and collated in NSSC; Searle (2014) Victoria Government (Mark Imhof, pers. Comm.)

EVI, Enhanced vegetation index; VPD, vapour pressure deficit; ESP, exchangeable sodium percent; EC, electrical conductivity; Cl, chloride; BD, bulk density; ACLUMP, Australian collaborative land and management program; ABS, Australian Bureau of Statistics; CSIRO, Commonwealth Scientific and Industrial Research Organisation; USGS, United States Geological Survey; MODIS, MODerate Resolution Imaging Spectrometer; SLA, Statistical Local Area ; SA2, Statistical Area Level 2; SILO, Scientific Information for Land Owners; NSSC, National Soil Site Collation

*Method details: **ten steps to estimate economic impacts of soil constraints***

Step (i): Define active-cropping areas

The ACLUMP land use map (Figure 1; http://www.agriculture.gov.au/abares/aclump) provided the basis for the area of cropping. The dryland and irrigated cropping and cereal cropping classes of the map represent areas that have at some point in recent history been observed (or otherwise inferred) to have been cropped. As such, we regard this map to effectively represent a maximum extent for the area sown with wheat for any given cropping season, and we refer to it here as the 'potential cropping area'. For example, the total area over Australia falling into classes 3.3.0 (dryland cropping), 3.3.1 (dryland cereal cropping), 4.3.0 (irrigated cropping), and 4.3.1 (irrigated cereal cropping) was 38 million ha (ACLUMP), while the total area cropped with wheat in any single season averaged 13 million ha over the growing seasons from 1999–2012 (ABS). This variance represents the practice of crop rotations and seasonal variability that determines the actual area of wheat grown in any year.

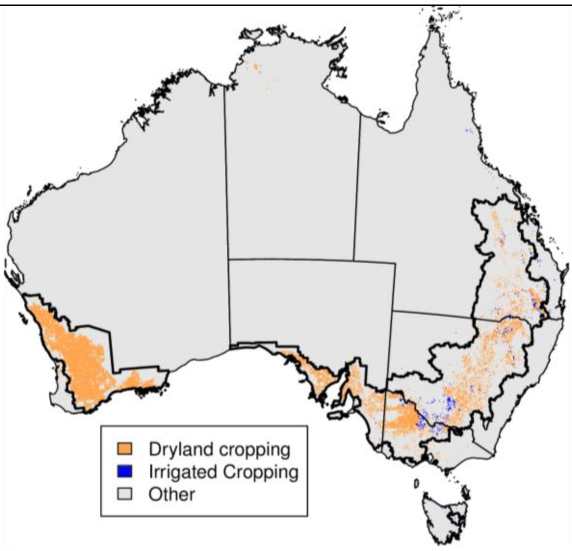


Figure 1 ACLUMP land use map with three categories. An approximate outline of the cropping region is superimposed, based on SA2 boundaries

The yield data that we will use in this work are available as average yields over SLAs (see Table 1 and Figure 2). Note that these differ slightly from the final required reporting level of SA2. More precisely, the datum for any given SLA and growing season is defined as an average yield over the area within that SLA that was cropped in that given season (which we refer to as the 'active cropping area' of an SLA for growing season *y*). We therefore consider the following approach to predict the active cropping areas of SLAs based on the ACLUMP map, ABS data on total area sown to wheat, and remote-sensing time-series data.

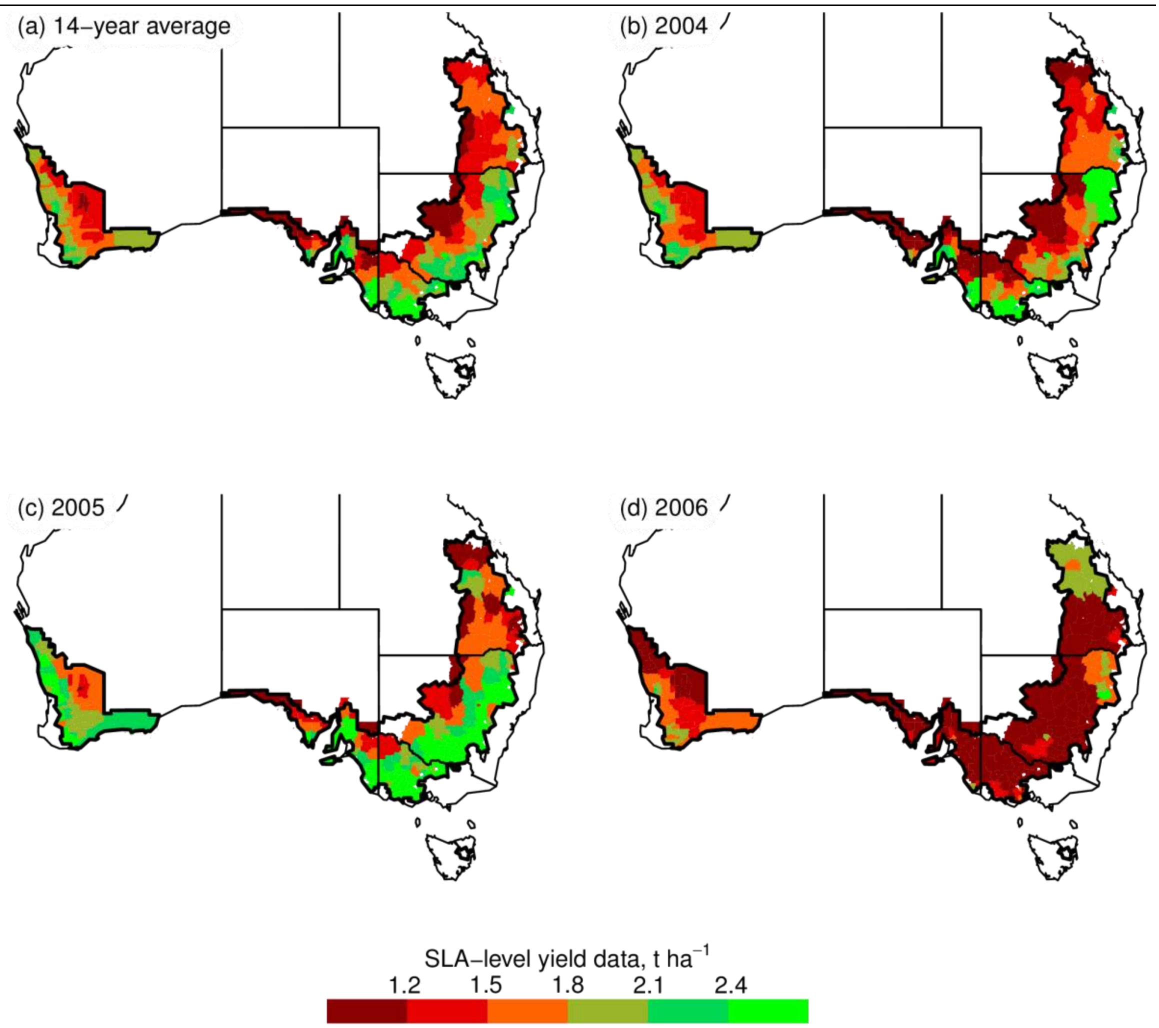


Figure 2 SLA-level yield data for (a) the 14-year average and (b), (c) and (d) three illustrative years, 2004, 2005 and 2006, respectively

The remote-sensing time-series data that we use come from Landsat 5-TM and 7-ETM+ (30-m pixels) and MODIS MOD13Q1 (250-m pixels), see Table 1. They comprise data on the enhanced vegetation index (EVI), a vegetation index similar to the normalized difference vegetation index (NDVI) that is more robust at high biomass and to soil interference (Huete et al., 1997). The Landsat data were processed for surface reflectance using the method of Flood et al. (2013). For any 30-m pixel within the potential cropping area and any growing season, a single time series was created by merging the Landsat data that fell within the period May–November with those from MODIS that covered the same pixel for the same period. We note that a single MODIS pixel could therefore provide data for a number of time series.

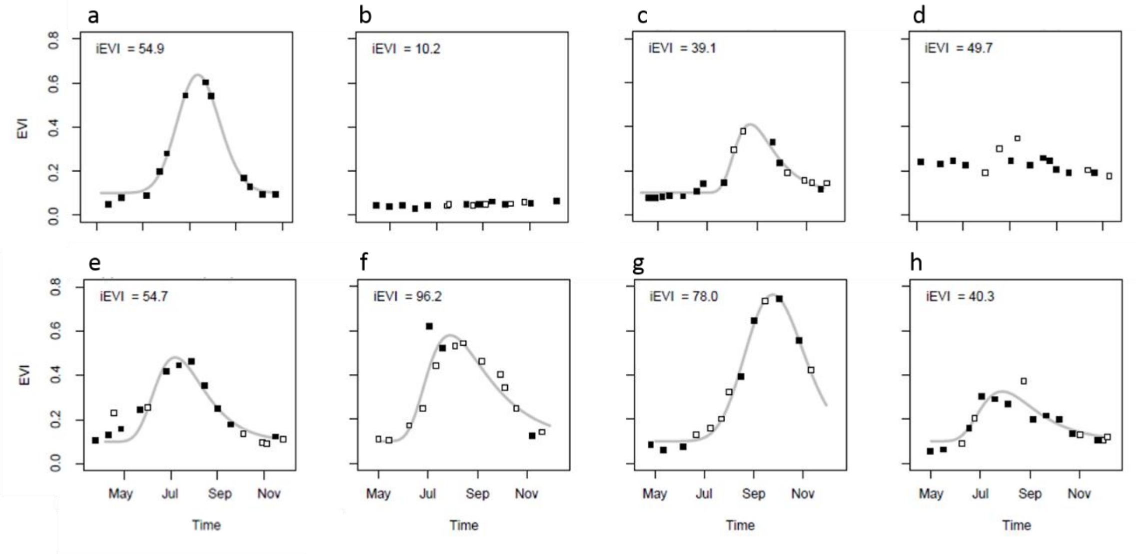


Figure 3 EVI time series for randomly selected location-year combinations. Solid symbols are Landsat observations, open symbols are MODIS. A grey line (line of best fit) indicates significant 'crop-like' growth, which we assume to be wheat. The time-integrated EVI, iEVI, is also given in each case. The selected pixels fall in the following SLAs and years: (a) Buloke (S) – North, 1999; (b) Belyando (S), 2001; (c) Ceduna (DC), 2004; (d) Lake Grace (S), 2006; (e) Wakefield (DC), 2007; (f) Perenjori (S), 2008; (g) Wambo (S), 2010; (h) Moree Plains (A), 2012

Each of these time series was then analysed for evidence of a bell-shaped curve (Werker-Jaggard growth; Werker and Jaggard, 1997). If such evidence existed (based on the Akaike Information Criterion, AIC; Akaike, 1973), the pixel was given an initial classification of 'cropped', the assumption being that the pixel was cropped with wheat[2]. This analysis was repeated for all 30-m pixels within the ACLUMP 'potential cropping area', and for all growing seasons from 1999–2012. However, this treats all pixels as independent, whereas crops are grown (or not) on a paddock basis. Furthermore, the area of a state predicted to have been sown in any season might be vastly different from the ABS data for the same (Table 2). Therefore, we used an object-oriented classification algorithm (Bunting et al., 2014) to define paddock boundaries across Australia, and then applied the following algorithm to classify paddocks (i.e. all pixels within a paddock) as cropped or not cropped, so as to as closely as possible match the predicted and observed (Table 2) total areas sown to wheat for each of the three disjoint zones (Table 2) and each of the 14 growing seasons. The steps of the algorithm were:

i. Set an initial cut-off value as w = 50 % (see Step iii.)
ii. For each paddock, p, in zone S and growing season y, calculate the percentage of the paddock's pixels whose initial classification was 'cropped', $P_c(p, S, y)$
iii. For each paddock, p, in zone S and growing season y, if the paddock's percentage of initially-classified cropping pixels, $P_c(p, S, y)$, is greater than w, then classify the entire paddock as 'cropped', else classify the entire paddock as 'not cropped'
iv. Sum up the area of land in zone S and growing season y classified as 'cropped', $\widehat{A_c}(S, y; w)$, based on the classifications of paddocks in Step iii.

[2] Since we are unable to distinguish between different types of winter crops, some of the land classified as cropped in the current work will have been cropped with something other than wheat. However, as wheat is by far the dominant winter crop in terms of area sown (in the three years 2005–2006, 2006–2007 and 2007–2008 an average of 12.3 M ha was sown to wheat, compared with 4.5 M ha for barley and 1.1 M ha for canola; ABS), the errors should not be cause for concern, particularly when our ultimate objective is to report results at broad rather than local scale.

v. Compare $\widehat{A_c}(S\,,y\,;w)$ with the datum for the total area sown to wheat in zone *S*, growing season y, $A_c(S\,,y)$. If $\widehat{A_c}(S\,,y\,;w) > A_c(S\,,y)$, then the predicted area of cropping is too large, so increase w; if $\widehat{A_c}(S\,,y\,;w) < A_c(S\,,y)$, then the predicted area of cropping is too small, so decrease *w*

vi. Return to Step iii. and iterate until the predicted area of cropping for zone *S* and growing season *y* is sufficiently close to the actual area of cropping for *S* and *y*.

Table 2 ABS data on area cropped to wheat, (000 ha) for three disjoint zones (*S* in the above algorithm); note, Eastern Australia includes all of Australia except for Western Australia and Tasmania

Year	Western Australia	Tasmania	Eastern Australia
1998	4515	4	7023
1999	4556	6	7605
2000	4460	7	7675
2001	4350	6	7173
2002	4458	7	6705
2003	4917	8	8142
2004	5118	7	8274
2005	4753	8	7682
2006	4037	7	7754
2007	4258	7	8313
2008	4542	9	8980
2009	5006	7	8868
2010	4640	8	8854
2011	5156	7	8740

All steps in the above algorithm were repeated for each zone, *S*, and each growing season, *y*, resulting in active cropping maps for all growing seasons. We note here a limitation of the approach. It is possible that a soil constraint could render an entire paddock infertile, and have such an impact that none of its pixels show Werker-Jaggard-type growth. Such paddocks would be classified as not-cropped by our analysis approach, and any soil data from that paddock—possibly identifying the responsible constraint—would not contribute to our yield gap estimates. This could be an issue for soil salinity, which can destroy entirely the ability of land to produce.

To illustrate results from this step, Figure 4 shows the Waggamba (S) SLA in southern Queensland. Potential cropping land is depicted as either green or brown, the green showing areas of active cropping for three illustrative years, 2004 (Figure 4a), 2005 (Figure 4b) and 2006 (Figure 4c). The maps produced in this step are used in Step (ii) of the yield-gap algorithm to disaggregate areal yield data over the active cropping areas of SLAs. Accordingly, at the soil data locations we will have yield data from some—but not necessarily all—years.

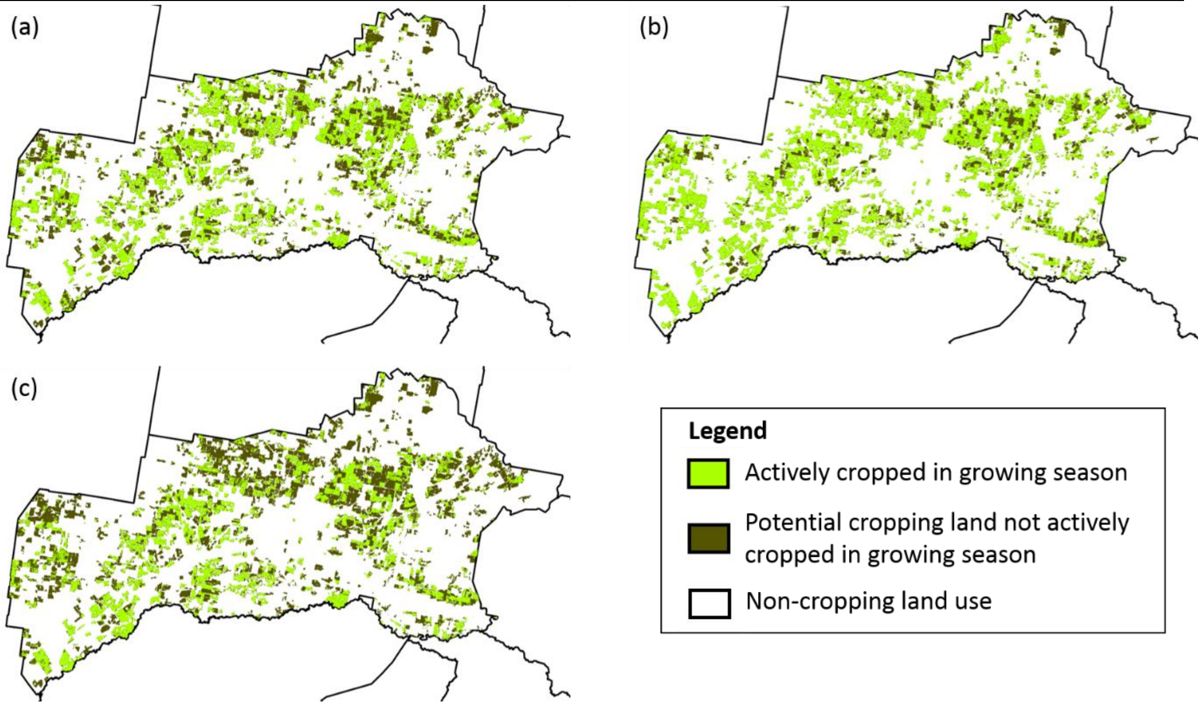


Figure 4 An illustration of the results from Step (i), defining the active cropping regions. The SLA Waggamba (S) in southern Queensland is shown for three illustrative years, (a) 2004, (b) 2005 and (c) 2006, with all potential cropping land (according to the ACLUMP land use map) coloured green or brown; green if actively cropped in that growing season, and brown otherwise.

Step (ii): Disaggregate yield to soil data locations

Our yield data for each growing season represent average yields over the active cropping areas of SLAs. In order to fit a model that represents yield as a function of soil, we require yield information at the spatial support of the soil data (soil cores, or effectively point support). We therefore harmonize the spatial support of the yield data to that of the soil data, through a process known as spatial disaggregation. Area-to-point (ATP) kriging (Kyriakidis, 2004) is a geostatistical method designed for exactly this purpose. Areal data ($\overline{y}$, our yield data as SLA averages) are defined as an average of the point-support variable (y, the yield at points within the SLAs) over spatial units (the active crop growing regions of SLAs), and this relationship can be used to write a statistical model (effectively a mixed-effects model; see Pinheiro and Bates, 2004) that links the areal-support data and point-support predictions. This framework allows a spatial trend for the underlying point-support variable, y, to be defined as a function of some other relevant mapped variable (i.e. one that is available on a fine-scale mapping grid). In our case, the time-integrated EVI, iEVI, provides us with this trend variable, and we consider a sigmoidal function of iEVI so that its extreme values do not give rise to nonsensical predictions of yield. Parameters of the statistical model were fitted by residual maximum likelihood (REML; Patterson and Thompson, 1971). Further details on the ATP-kriging method can be found in Kyriakidis (2004), and on the use of REML to estimate the model parameters in Orton et al. (2017). The result of this disaggregation procedure was a set of predicted yields at the locations of our soil data for each year the locations were deemed to have been cropped (see Step (i)).

Step (iii): Extract soil data and harmonize soil depths

The soil data we work with come from two sources. First, the National Soil Site Collation (NSSC), which contains data from over 280 000 soil profiles, derived from the soil data holdings of each state and territory, plus the Commonwealth (Searle, 2014). Second, data were obtained from Victoria Government

(Mark Imhof, pers. comm.) and used to supplement those from the NSSC. Data were extracted from both databases that fell within the potential cropping area only: any data within 30 m of land that was not classified as potential cropping (according to the ACLUMP map) were excluded. Data were extracted for the following soil properties: exchangeable sodium percentage (ESP), pH, electrical conductivity (EC), chloride concentration (Cl), bulk density, and soil texture components (sand, silt and clay). Measurement methods for each of the soil variables are described in Rayment and Lyons (2011), and we refer to their method codes throughout the following sections. The data included measurements of soil samples collected from soil horizons within the profiles, and others collected from fixed (i.e. determined prior to sampling) depth intervals. After describing the extraction of appropriate data to represent the impacts of soil constraints, we describe the harmonization of soil depths for these data.

Exchangeable sodium percentage: ESP

Soil sodicity—an excess of sodium ions in relation to other cations—can lead to soil dispersion and poor soil structure and thus have an impact on wheat yield. The ESP of soil provides an indicative measure of soil sodicity and dispersion. Although other measures, such as the sodium adsorption ratio (SAR; e.g. DERM, 2011), have also been proposed, we use the ESP due to the greater availability of data. ESP is calculated from the concentration of the sodium cations and the soil's cation exchange capacity (CEC). Soil Science Australia (2015) provide guidelines to determine the most appropriate method for measuring the cations and CEC, which depend on the pH and EC of the soil sample, and we follow the preferences outlined therein. This gave us ESP data from 2542 profiles within the wheat cropping land.

pH

In soils with a low pH, aluminium and manganese become more soluble and can inhibit wheat root growth (Tang et al., 2003a). Soil pH data were used to provide a measure of soil acidity. The two most commonly used methods for measuring soil pH are; in a 1:5 soil:water solution, method code 4A1, and in a 1:5 0.01 M calcium chloride solution, method code 4B1. We refer to pH data measured through method 4A1 as pH_W, and those measured through 4B1 as pH_{Ca}. The latter is usually preferred as it provides a more consistent measurement that is less affected by soil electrolyte concentration (Minasny et al., 2011). We therefore base our measure of soil acidity on pH_{Ca} data. However, there is a strong spatial clustering in these data, with Western Australian pH data being predominantly pH_{Ca} and the rest of Australia mainly pH_W. Calibration equations have been presented for conversion of pH between the two measurement methods. Minasny et al. (2011) present an equation (derived by fitting a neural network to a comprehensive dataset) to predict pH_{Ca} given measurements of pH_W and EC (EC, itself measured in 1:5 soil:water solution, method code 3A1). Also, Henderson and Bui (2002) used a spline to model pH_{Ca} as a function of pH_W, and present a look-up table summarizing the results. Whenever pH_{Ca} data are missing from our database we predict it (if possible) using the equation of Minasny et al. (2011) if both pH_W and EC data are available, or the look-up table of Henderson and Bui (2002) if just a pH_W measurement is available. After conversion of pH_W to pH_{Ca} data (where the latter were missing), our database consisted of data from 6513 profiles. From here on, we will work exclusively with the resulting pH_{Ca} data and will refer to these solely as pH for simplicity, only referring to the method of measurement where necessary to avoid confusion.

Electrical conductivity: EC

The EC provides an indication of soil salinity as a measure of the total concentration of salts (DERM, 2011). The EC data we work with are measured in 1:5 soil:water solution, method code 3A1, in units of dS m^{-1}. Across the cropping land of Australia, this gave us EC data from 2962 soil profiles.

Chloride concentration: Cl

The Cl concentration provides additional information to the EC measurement for interpretation of soil salinity (DERM, 2011). Our Cl data are measured in 1:5 soil:water extract, method code 5A2, measured in units of mg kg^{-1}. In a region of Western Australia, Cl data measured by this method were scarce, therefore the data were supplemented here with data reported under the code 5NR (data for which the laboratory method was for some reason not recorded), after checking for consistency with the 5A2 data. This gave us Cl data from 820 profiles.

Bulk density: BD

The bulk density provides information about soil compaction. It is most commonly measured in Australia by determining the weight of an intact soil core of known dimensions. Our BD data came from a total of 913 soil profiles across Australia. However, we note here that bulk density data commonly exhibit a large degree of local variability, due in part to machinery-induced compaction being a very localized process, and in part due to measurement difficulties (see e.g. Don et al., 2007; Allen et al., 2011). As a result of this local variability, and of the fewer bulk density data compared with those for the soil chemical constraints, we might encounter difficulties in modelling its effects based on data that are derived from satellite imagery.

Soil texture: sand and clay contents

The impact that any of the soil constraints we consider will have on wheat yield will depend on other soil characteristics, in particular soil texture. We therefore included sand and clay data, measured by a number of different methods (Plummet balance, Hydrometer method, Coventry and Fett pipette method), giving sand and clay data from 2224 and 2380 soil profiles, respectively. When we fit models representing yield as a function of climate and soil, we will use only laboratory-measured soil texture data. However, when it comes to applying these models in order to calculate yield gaps due to soil constraints, we will allow data derived from hand-texture assessments to also be used.

Harmonization of soil depths

All the above soil properties are measurements on samples collected over soil horizons or over fixed depth intervals within soil profiles. These sampled depths therefore vary between profiles. We require information for some consistent depths that will allow us to model the effects that, for instance, soil acidity in the 0–10 cm soil depth has on yield. We harmonize our soil data to three depths: A: 0–10 cm, B: 10–50 cm and C: 50–200 cm. (We note here that depths A, B and C should not be confused with soil horizons.) We do so by thickness-weighted averaging, as follows. Suppose a soil profile contains data for m sampled depth intervals with upper and lower limits $u_i, i = 1, \dots, m$ and $l_i, i = 1, \dots, m$, respectively, with corresponding soil property measurements, $z_i, i = 1, \dots, m$. To calculate the thickness-weighted average Z for a depth interval with limits u_0 and l_0 (i.e. one of the intervals A, B or C defined above), we first calculate the length of intersection, L_i, of each data depth interval, $[u_i, l_i]$, with the prediction depth interval, $[u_0, l_0]$. The weights were calculated by standardizing the L_i so that they sum to one:

$$w_i = \frac{L_i}{\sum_{i=1}^{m} L_i} \tag{3}$$

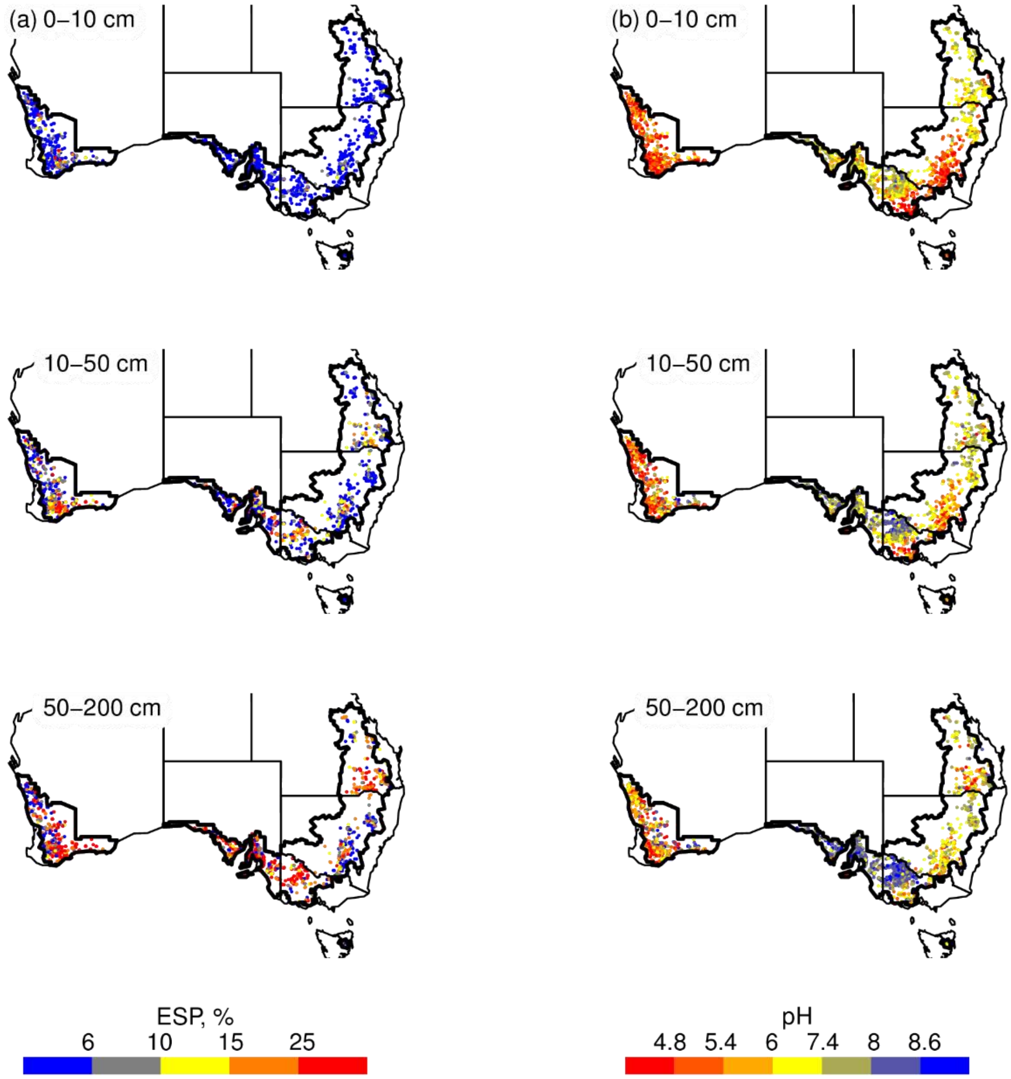
and the thickness-weighted average of Z for this profile is then:

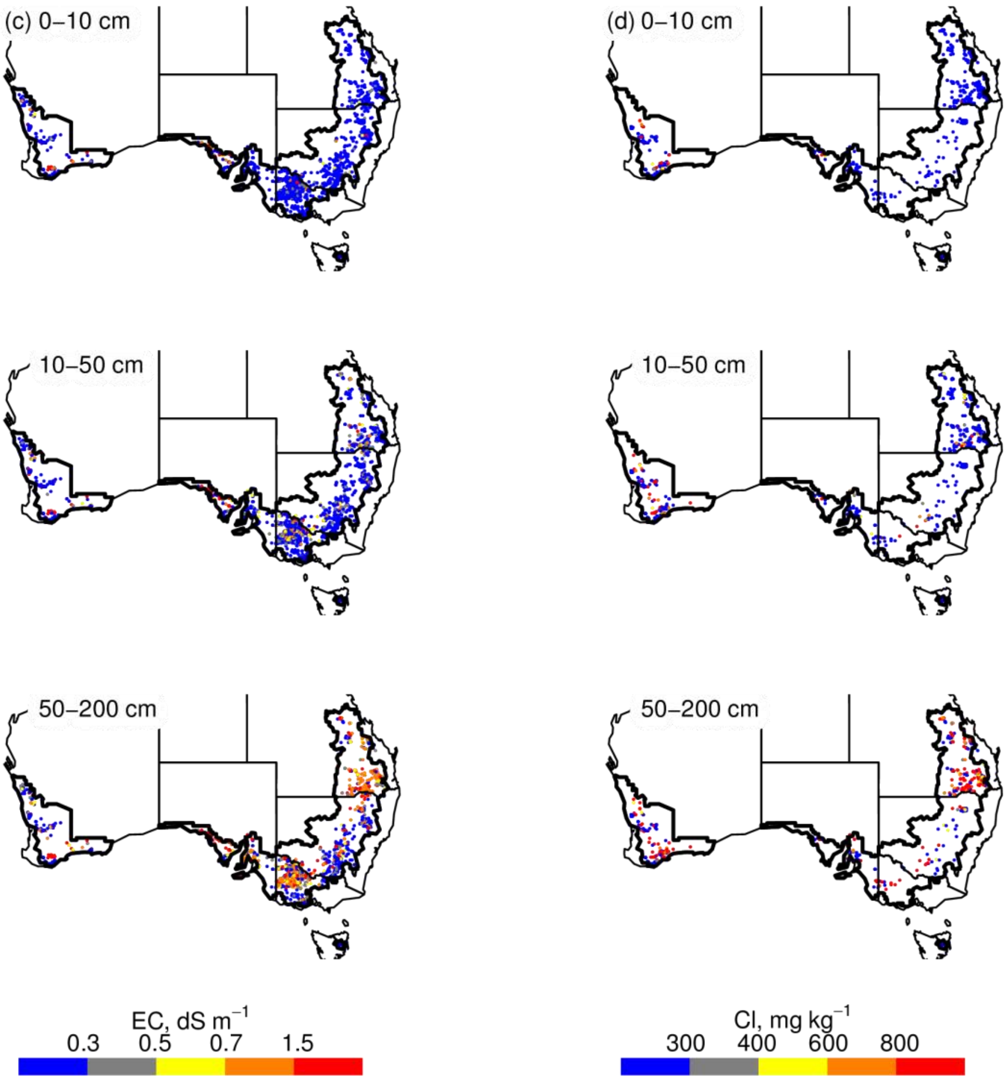
$$\hat{z}_0 = \sum_{i=1}^{m} w_i z_i \tag{4}$$

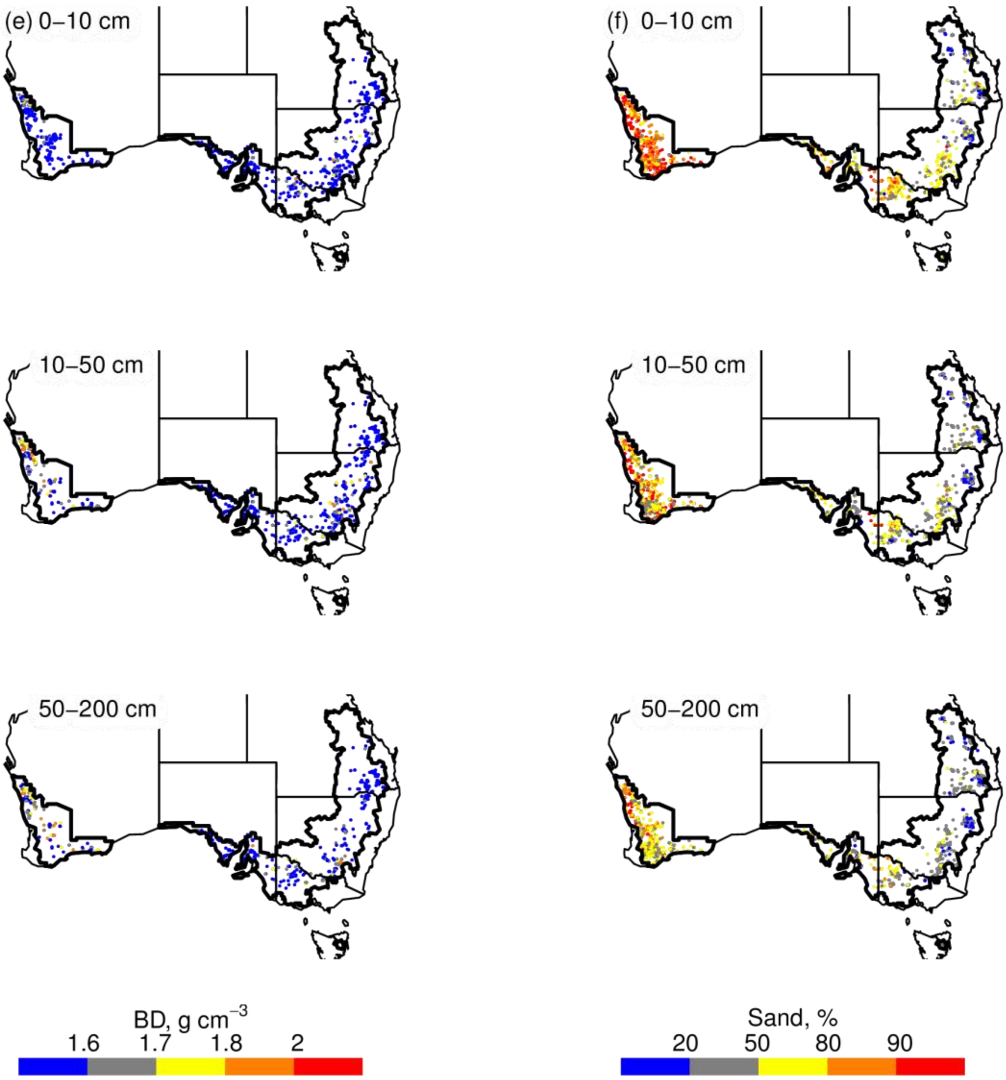
Table 3 summarizes the number of data for pH, ESP, EC, Cl, BD, sand, silt and clay. Since we will fit models that represent the impact of each of the soil constraints while accounting for soil texture, the bracketed numbers give the number of locations for which data for the stated soil property along with data for both sand and clay were available. Figure 5 shows the spatial distributions of the depth-harmonized soil data.

Table 3 Numbers of locations providing data for each soil variable at the three separate soil depths (A: 0–10 cm, B: 10–50 cm, C: 50–200 cm) and for all three soil depths together. Numbers in parentheses are the numbers of locations where both sand and clay data were also available

depth	ESP	pH	EC	Cl	BD	Sand	Clay
A	1960 (1398)	6265 (1548)	2835 (924)	761 (402)	830 (366)	1880 (1880)	2035 (1880)
B	2251 (1562)	5335 (1730)	2879 (962)	761 (382)	802 (390)	2029 (2026)	2182 (2026)
C	1931 (1295)	3874 (1402)	2506 (849)	741 (359)	637 (326)	1679 (1678)	1834 (1678)
A+B+C	1483 (1015)	3658 (1154)	2364 (733)	679 (335)	553 (243)	1418 (1417)	1572 (1417)




(c) 0–10 cm
10–50 cm
50–200 cm
(d) 0–10 cm
10–50 cm
50–200 cm
EC, dS m⁻¹
0.3 0.5 0.7 1.5
Cl, mg kg⁻¹
300 400 600 800



(e) 0–10 cm
10–50 cm
50–200 cm
(f) 0–10 cm
10–50 cm
50–200 cm
BD, g cm⁻³
1.6
1.7
1.8
2
Sand, %
20
50
80
90

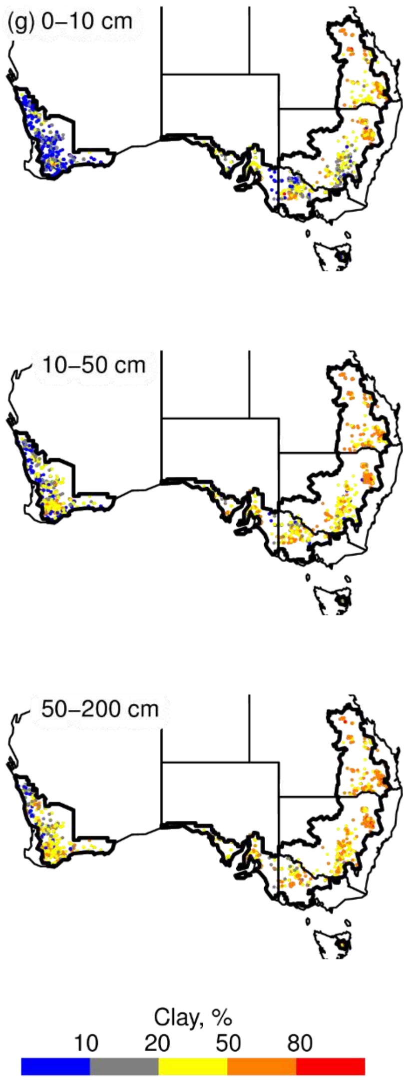


Figure 5 Depth-harmonized soil data for (a) ESP, (b) pH, (c) EC, (d) Cl, (e) BD, (f) sand and (g) clay, for the three depths, A = 0–10 cm (upper panels), B = 10–50 cm (middle panels) and C = 50–200 cm (lower panels)

Step (iv): Define optimal ranges for soil constraints

For each of the four soil constraints under consideration we define optimal ranges of values; that is, values for which we would not expect any detrimental effects of the constraints on wheat yield. We also define the type of behaviour we would expect of yield as the soil constraint departs from its optimum range. This will allow us to fit models that represent sensible responses of wheat yield to soil properties. We consider each soil constraint in turn.

Soil sodicity: ESP

Previous work on the effects of soil sodicity (DNRM, 2007) has considered potential sodicity constraints at values of ESP above 6 %, irrespective of soil depth, and we therefore define our optimum ESP as 0–6 %; these critical values of ESP for the effects of sodicity, $ESPCrit_{Sdcty}[d]$, are given in Table 4. Since our ESP data have a large skew we work with the log-transformed data. That is, our optimum values for the log-transformed variable, *lnESP*, are anything less than 1.8 (= ln 6). Yield penalties will be incurred for larger values of *lnESP*; if a yield penalty of 0.1 t ha^{-1} were to be modelled for a soil with $ESP[\mathrm{A}]$= 16 % (i.e. $lnESP[\mathrm{A}]$ = 2.8, one unit above the critical value for $lnESP[A]$), then a yield penalty of 0.2 t ha^{-1} would be modelled for a soil with $ESP[\mathrm{A}]$ = 44 % (i.e. $lnESP[\mathrm{A}]$= 3.8, two units above the critical value).

$$sdcty[d] = \begin{cases} lnESP[d] - \ln ESPCrit_{Sdcty}[d] & \text{if } ESP[d] > ESPCrit_{Sdcty}[d] \\ 0 & \text{otherwise} \end{cases}$$

For shorthand, we will also refer to the variables with a subscript A, B or C indicating the depth, for instance $sdcty_A$ for $sdcty[d = \mathrm{A}]$; we will follow a similar convention for all the soil constraints.

Table 4 Summary of critical values for all soil constraints and for the three depths, *d*; A: 0–10 cm, B: 10–50 cm, C: 50–200 cm.

d	$ESPCrit_{Sdcty}[d]$	$pHCrit_{Acdty}[d]$	$pHCrit_{Alkty}[d]$	$ECCrit_{Slnty}[d]$	$ClCrit_{Slnty}[d]$	$BDCrit_{Cmpctn}[d]$
A	6.0	6.0	7.4	0.3	300	1.6
B	6.0	4.8	7.4	0.7	600	1.6
C	6.0	4.8	7.4	0.7	600	1.6

Soil acidity: pH

Work on soil acidity and the effects of liming in Western Australia (Gazey and Davies, 2009) has considered soil pH target values of 5.5 in the topsoil and 4.8 in the subsurface. Australia-wide, soils with topsoil pH less than 6.0 have been considered to be slightly acidic (AACM International, 1995). The effects of aluminium toxicity in the subsoil are minimized if the subsoil pH is above 4.8, while the higher topsoil pH target values allow alkalinity, through applied lime, to migrate to lower soil layers and neutralize acidity down the profile (Gazey and Davies, 2009). At the other end of the pH scale, alkalinity can result in detrimental effects on wheat yield; DNRM (2007) considered that yield penalties occur at pH > 7.4 (converted from pH_W > 8.0; Henderson and Bui, 2002). In this work, we consider the optimum range for pH in the 0–10 cm soil depth as 6.0–7.4, and in the 10–50 and 50–200 cm soil depths as 4.8–7.4. We denote the variables representing these critical values of pH for acidity and alkalinity as $pHCrit_{Acdty}[d]$ and $pHCrit_{Alkty}[d]$, respectively, where d represents the soil depths A, B or C. Their definitions are summarized in Table 4. If the soil pH falls outside these ranges, then we expect yield penalties to occur linearly: that is, if a yield penalty of 0.1 t ha^{-1} were to be modelled for a soil with a pH of 5.0 in the 0–10 cm depth (i.e. 1 pH unit below the critical value of 6.0), then a yield penalty of 0.2 t ha^{-1} would be modelled for a soil with similar properties except for a pH of 4.0 in the 0–10 cm depth (i.e. 2 pH units below the critical value). The assumption of linear effects of acidity below the critical values is justified on the basis of the level of our data, which are not sufficiently detailed to allow more complex modelling assumptions (our yield data, as described in Step (ii), come from many different wheat varieties of differing aluminium tolerance, all farmed under unknown management conditions).

In order to fit a model that represents the above information, we define the following variables for the degrees of acidity and alkalinity, respectively:

$$acdty[d] = \begin{cases} pHCrit_{Acdty}[d] - pH[d] & \text{if } pH[d] < pHCrit_{Acdty}[d] \\ 0 & \text{otherwise} \end{cases}$$

$$alkty[d] = \begin{cases} pH[d] - pHCrit_{Alkty}[d] & \text{if } pH[d] > pHCrit_{Alkty}[d] \\ 0 & \text{otherwise} \end{cases}$$

where d again represents the soil depths A, B or C; the critical values are defined in Table 4, and $pH[d]$ is the pH data for depth d.

Soil salinity: EC and Cl

On its own, EC measures total salts, some of which (e.g. gypsum) might not result in negative impacts on crop yields. We therefore also consider the chloride (Cl) concentration to provide a measure of the Cl component of soil salinity. We base our definition of the effects of salinity on the decision tree presented in DNRM (2007), which considered potential salinity constraints if the topsoil (0–10 cm) EC was larger than 0.3 dS m^{-1} or the subsoil (below 10 cm) EC was larger than 0.7 dS m^{-1}. If these critical values were exceeded, topsoil Cl concentrations of greater than 300 mg kg^{-1} and subsoil concentrations of greater than 600 mg kg^{-1} were considered to be potentially constraining to wheat growth (see Table 4). We model potential penalties that increase with the log of the Cl concentration (due to the large skew of Cl): if the log-transformed variable, *lnCl*, exceeds 5.7 (i.e. *Cl* exceeds 300 mg kg^{-1}) for 0–10 cm or 6.4 (i.e. *Cl* exceeds 600 mg kg^{-1}) for 10–50 and 50–200 cm, then yield penalties are potentially incurred linearly with *lnCl*. For example, if a yield penalty of 0.1 t ha^{-1} were to be modelled for a soil with $Cl[\mathrm{C}]$ = 663 mg kg^{-1} (i.e. $lnCl[\mathrm{C}]$ = 6.5, 0.1 units above the subsoil Cl limit of $lnCl[\mathrm{C}]$ = 6.4), then a penalty of 0.2 t ha^{-1} would be modelled for a similar soil with subsoil $Cl[\mathrm{C}]$ = 733 mg kg^{-1} (i.e. $lnCl[\mathrm{C}]$ = 6.6, 0.2 units above the subsoil Cl limit). To represent this information on soil salinity, we define the variable ‘degree of Salinity’, *slnty*, for the three soil depths, $d = \mathrm{A}, \mathrm{B}, \text{or } \mathrm{C}$, as:

$$slnty[d] =$$

$$\begin{cases} lnCl[d] - \ln ClCrit_{Slnty}[d] & \text{if} \begin{cases} \left(Cl[d] \geq ClCrit_{Slnty}[d]\right) \\ \text{and} \\ \left(EC[d] \geq ECCrit_{Slnty}[d] \text{or } EC[d] \text{ unknown}\right) \end{cases} \\ 0 & \text{if } EC[d] < ECCrit_{Slnty}[d] \end{cases}$$

Soil compaction: BD

Soil compaction has a detrimental effect on root growth and although its impact can depend on the soil type, in general it tends to occur when the bulk density exceeds 1.6 g cm^{-3} (McKenzie et al., 2004). We therefore consider potential effects of compaction for bulk density over the critical value of $BDCrit_{Cmpctn}[d]$ = 1.6, for $d = \mathrm{A}, \mathrm{B}, \text{or } \mathrm{C}$ (see Table 4), and our definition of the degree of compaction is:

$$cmpctn[d] = \begin{cases} BD[d] - BDCrit_{Cmpctn}[d] & \text{if } BD[d] > BDCrit_{Cmpctn}[d] \\ 0 & \text{otherwise} \end{cases}$$

Step(v): Pre-process climate data

The most important determinant of plant growth is the availability of water, which in turn is largely governed by climate through rainfall and evaporation (Hughes et al., 2011). Therefore, strong seasonal and broad-scale locational patterns in yield data arise as a result of temporal and spatial variability in climate. We represent this climate variability through vapour pressure deficit (VPD), the difference between the actual vapour pressure and the vapour pressure under saturated conditions (Jeffrey et al., 2001). Daily vapour pressure and maximum and minimum temperatures were extracted from the SILO 5-km gridded dataset (https://www.longpaddock.qld.gov.au/silo/index.html), and used to calculate daily values of VPD (Tetens, 1930). These values were averaged over the months of September and October for each year to give a VPD value for each 5-km grid cell in each growing season (1999–2012). September and October were chosen to represent the period when moisture availability is most important for winter wheat growth over the broad range of Australia’s wheat-growing regions, though we acknowledge that any single period could not be so for the entire country. The calculated values were assumed to be representative of the VPD within each 5-km grid cell.

These VPD values exhibit variability in both space and time, and result in yield differences in both space and time. In order to investigate these spatial and temporal effects of climate, we therefore split the VPD into two variables: a spatial (long-term average) component (which we will denote VPD_S), and a temporal (yearly departure from the long-term average) component (VPD_T). Figure 6a shows the VPD_S component, and Figure 6b, 6c and 6d the VPD_T component for three exemplary years, 2004 (generally moderate), 2005 (generally wetter than average) and 2006 (generally drier than average), respectively. The small VPD_S values in the south (Figure 6a) represent the higher rainfall areas, where conditions are generally more favourable for plant growth.

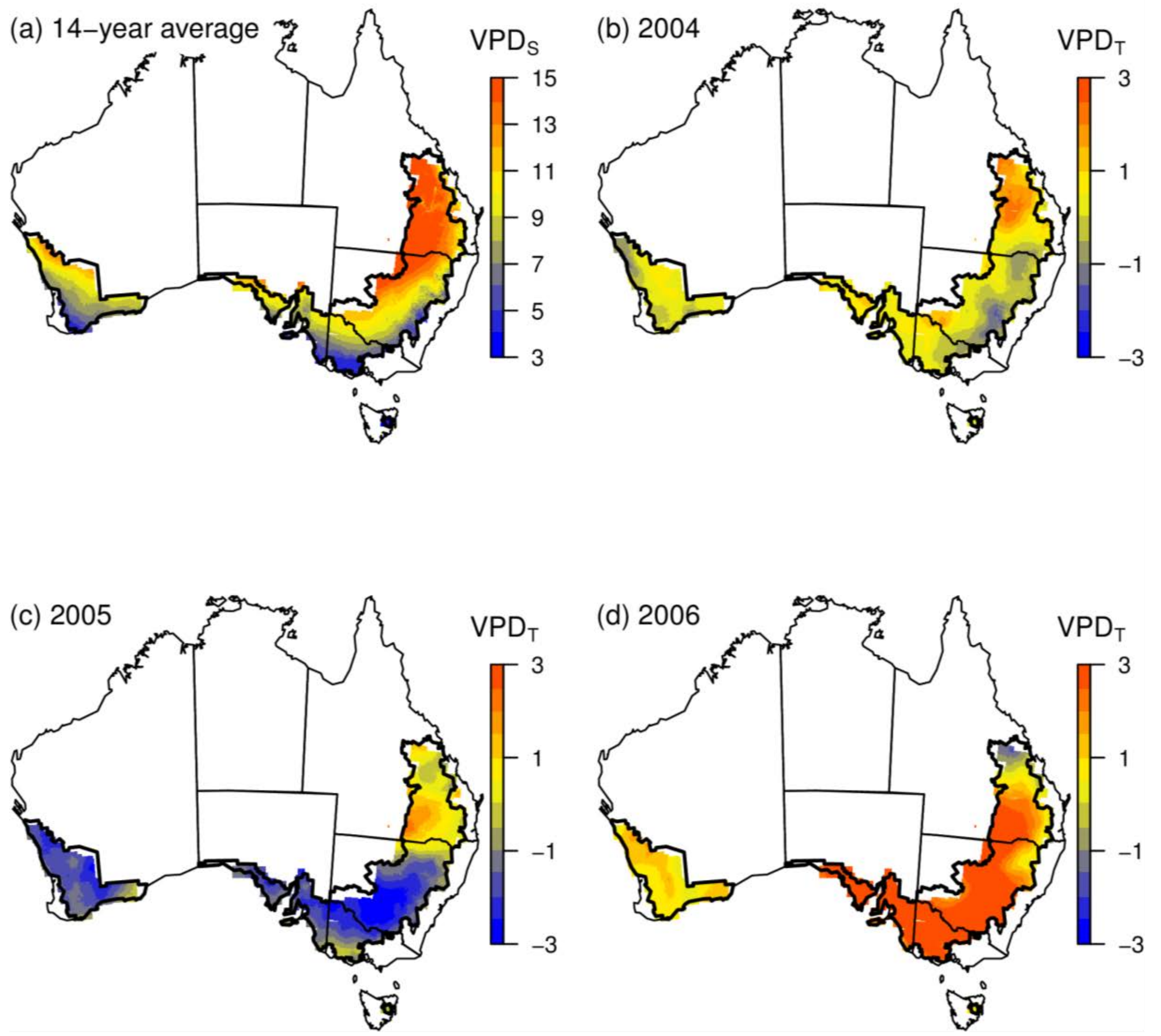


Figure 6 The components of VPD, in units of kPa; (a) spatial (long-term average) component, VPD_S; (b), (c) and (d) temporal (yearly departure from long-term average) component, VPD_T, for 2004, 2005 and 2006, respectively

Step (vi): Fit a model that represents yield as a function of soil constraints

Steps (i)–(v) have produced a dataset that consists of yield, climate and soil variables (in the form of degrees of each constraint) on a similar spatial support; in total, this gave 54 744 yield data at the soil data locations over 14 years, though not all location-year combinations presented data for all variables (e.g. if a location was deemed in Step (ii) to not be cropped in a certain year, or if at a certain location the soil's ESP was not measured). In addition, we included in our dataset an indicator variable representing 'irrigated' or 'dryland' cropping based on the classifications of the ACLUMP map. The dataset thus includes very limited management information, due to the lack of management data at the specific sites used for fitting models across Australia. It is therefore assumed that other management factors contributed random errors to the following models. The modelling approach that we pursue is based on a combination of a machine-learning algorithm, Cubist, and a linear mixed-effects statistical model; we utilize output from the former and apply the latter framework to account for the structure of the dataset when fitting parameters. We now describe the approach in more detail.

For each of the four soil constraints in turn (sodicity, acidity, salinity and compaction), Cubist models (Quinlan, 1992) were fitted to the data. Cubist is a machine-learning software tool that can be used to fit 'model-tree' regression models; a set of rules that define subsets of the data and a corresponding set of multiple regression models that apply under the conditions of each rule. The Cubist model rules allow interactions between predictor variables and non-linear effects to be modelled, and the combination of these rules with multiple regression models allows linear effects to be modelled where appropriate. Although the prediction model produced by the Cubist software can represent more complex features (e.g. the use of 'committees' to model residuals from successive model trees and thereby reduce prediction errors), in this work we use the basic model-tree form.

Cubist models were initially fitted with five rules, though if the resulting model used the soil constraint under consideration in its rule definitions, or if a rule was defined based on a minimal range of a variable (e.g. a rule defined as $8.1<VPD_S<8.2$), then a model with one fewer rule was fitted. The former of these conditions was imposed to ensure that a definition of a yield gap would come under one rule of the model (i.e. changing the value of the soil constraint from its actual one to its optimum would not result in a change of the applicable model rule and associated regression model). The latter condition was used to remove highly-specific and non-general rules. Missing data were handled by an iterative model-fitting procedure, as outlined in Appendix A. We will write a general form for the resulting Cubist prediction model as:

$$\hat{\mathbf{y}} = \mathbf{f}_c(\mathbf{X}, \hat{\boldsymbol{\beta}}) \qquad (5)$$

where $\mathbf{X}$ is the matrix containing the covariates for which $\mathbf{y}$ is to be predicted, $\hat{\boldsymbol{\beta}}$ the multiple regression parameters fitted by Cubist, and $\mathbf{f}_c$ the Cubist function that combines these to give the predicted values, $\hat{\mathbf{y}}$. The four resulting Cubist models are presented in Appendix B.

Our dataset contains multiple data from the same spatial location (i.e. yield data for multiple years) and multiple data from the same year, and we would expect these data to be more similar than those at different locations or years. To account for this correlation structure, we treat the Cubist software as a variable-creation tool, and refit parameters of the collection of multiple regression models ($\hat{\boldsymbol{\beta}}$ in Equation 5) under the framework of a linear mixed model (LMM; Pinheiro and Bates, 2004). Pringle et al. (2016) applied a similar approach to fit a regression tree model in the framework of a LMM and thereby account for spatial correlation while fitting its parameters, and we refer to them and to Appendix A here for further details.

The refitted Cubist LMM was inspected for any parameters representing positive effects of soil constraints, and any such parameters removed. For instance, when the Cubist model for the effects of sodicity was refitted as a LMM, it included a parameter in Rule 4 that represented an increase in yield with increasing sodicity at depth 50–200 cm; this parameter was therefore removed from Rule 4 and the model refitted. The reduced model was then refitted, and the process iterated until no parameters representing positive effects of soil constraints remained in the model. The refitted parameters of the four Cubist models are given in Appendix B. Comparing the refitted and original multiple regression parameters that make up the Cubist model reveals some differences, and hints at the worth of accounting for the structure of the dataset through the LMM.

Step (vii): Apply the model at soil data locations to calculate Y_{ac}, Y_{oc} and Y_{gc}

The models fitted as described in Step (vii) can be used to predict yield as a function of climate and soil through Equation 5, using the values of $\hat{\boldsymbol{\beta}}$ that were refitted in the framework of the LMM. With **X** taken to be the actual values of the soil and climate variables at the data locations, these predicted values give us $Y_{ac}(\mathbf{x}_\mathrm{S}, \mathbf{t})$; the $\mathbf{x}_\mathrm{S}$ here denotes that these predictions are at the set of n_c locations with the data necessary to apply the Cubist model for all 14 years, and the $\mathbf{t}$ that they are for each of the 14 years from 1999 to 2012. To calculate $Y_{oc}(\mathbf{x}_\mathrm{S}, \mathbf{t})$, the columns of **X** representing the soil constraint are set to their optimum values (as defined in Step (iv)), and predictions calculated again through Equation 5. Finally, to calculate $Y_{gc}(\mathbf{x}_\mathrm{S}, \mathbf{t})$, the yield gap due to soil constraint *c* at the soil data locations for each of the 14 years, the values of $Y_{ac}(\mathbf{x}_\mathrm{S}, \mathbf{t})$ and $Y_{oc}(\mathbf{x}_\mathrm{S}, \mathbf{t})$ are plugged into Equation 2. Our interest lies in the average yield gap over all years. Therefore for each soil data location, the 14-year average yield gap due to constraint *c* was computed, which we will denote by $Y_{gc}(\mathbf{x}_\mathrm{S})$, the absence of the argument $\mathbf{t}$ distinguishing this temporally-averaged yield gap estimate from the point-in-time estimates $Y_{gc}(\mathbf{x}_\mathrm{S}, \mathbf{t})$.

Step (viii): Interpolate values to 1-km grid over entire cropping area

The above procedure produced estimates of the average yield gap due to soil constraint *c*, at the soil data locations. The soil data locations however do not provide uniform coverage of the wheat-growing area, therefore some form of interpolation was necessary to account for the uneven spread of yield gap estimates. The 14-year average yield gaps, $Y_{gc}(\mathbf{x}_\mathrm{S})$, were interpolated to a 1-km grid by kriging, with the aid of a soil order map (ASRIS, http://www.asris.csiro.au/) to give the spatial trend. To account for the generally larger uncertainty associated with larger yield gap estimates, the interpolation was undertaken in a lognormal setting (with any yield gap estimates less than 0.01 being set to 0.01 to permit the log-transform for all data). Parameters for the statistical model behind the interpolation (a LMM) were fitted by REML; see Webster and Oliver (2007) for more details on geostatistical methods. The result is a set of yield gap estimates at the grid locations, which we denote by $Y_{gc}(\mathbf{x}_\mathrm{k})$, $\mathbf{x}_\mathrm{k}$ representing the interpolation grid locations.

In addition to mapping the yield gap, we also map the area affected by each soil constraint. To do so, we convert the yield gap estimates at the data locations, $Y_{gc}(\mathbf{x}_\mathrm{S})$, into an indicator variable, $P_{gc}(\mathbf{x}_\mathrm{S})$, which takes the value 1 if the average yield gap at a data location is greater than 0 and the value 0 otherwise. We then apply geostatistical interpolation to this indicator variable, again with the soil order map giving the spatial trend. The result is a map of values, $P_{gc}(\mathbf{x}_\mathrm{k})$, lying between 0 and 1, each representing the probability that the grid location was subject to some form of diminished yield as a result of the soil constraint *c*.

Step (ix): Aggregate predictions to SA2 level

For presentation of results and for economic analysis, the predictions $Y_{gc}(\mathbf{x}_\mathrm{k})$ and $P_{gc}(\mathbf{x}_\mathrm{k})$ were aggregated from the mapping grid to SA2 level. We denote by $Y_{gc}(r)$ and $P_{gc}(r)$ the average yield gap due to and probability of being affected by, respectively, soil constraint *c* in an SA2 indexed by *r*. We denote by $Y_{gc}(r)$ the average yield gap due to soil constraint *c* in an SA2 indexed by *r*. Similarly, $P_{gc}(r)$ will denote the average probability in SA2 *r* of cropping land being affected by soil constraint *c*; an estimate of the area of land affected is thus $P_{gc}(r)A(r)$, where $A(r)$ is the area of the SA2 used for cropping.

Step(x): Apply economic analysis at SA2 level

We used a constraints analysis approach to estimate the forgone value of farm value added (farm surplus) to generate information on the economic impact of the soil constraints at the SA2 level (Mallawaarachchi et al., 1996). Estimation involved the following steps (see Table 5 for the figures):

a. Estimates of the gross value of forgone yield were calculated by multiplying $Y_{gc}(r)$ by the 2016 average delivered price for wheat, $260 per tonne (Ashton et al., 2016).
b. The marginal cost, the additional cost of producing a tonne of wheat, under regular management practices was estimated from a gross margin model (Rural Solutions SA, 2016), under the assumptions of progressively increasing yield, and the related costs for standard inputs, handling and levies, etc.
c. Then the average treatment costs for ameliorating each soil constraint, soil sodicity and acidity, were estimated separately using indicative costs (Upjohn et al., 2005; Petersen, 2015). For each constraint, these two components (the marginal cost of increased production and the cost of generating this increased production through soil amelioration) were summed to give the total marginal costs, per tonne of wheat, which were used to represent the total costs of amelioration of a specific constraint.
d. To determine the net value of forgone yield for each SA2, the cost of amelioration estimated at Step c was deducted from the estimated gross value of forgone yield from Step a.

The method thus accounts for both the costs of inputs used to generate the additional yield after amelioration and the costs of undertaking ameliorative action to alleviate a known constraint. Given the variable nature of the response to management practices for specific soil constraints at a given location, the estimates were made using average responses and average costs to generate indicative estimates of benefits.

Table 5 Key parameters used in economic assessment

Parameter	Unit ($)	Value
Average Wheat Price	per tonne (t), delivered	260
Marginal cost of wheat	per tonne, assuming an average yield of 2 t ha^{-1}	28.14
Treatment costs		
Gypsum for sodicity		
Gypsum rate	Tonnes per ha	2.5
Frequency of application	1 in 12 years, split annually	0.8
Price of gypsum	per tonne	35
Transport	per tonne	19
Application	per tonne	19
(b) Lime for acidity		
Liming rate	Tonnes per ha	2.8
Frequency of application	1 in 12 years, split annually	0.8
Price of lime	per tonne	32
Transport	per tonne	20
Application	per tonne	9

The approach draws on previous work (Mallawaarachchi et al., 1996; Petersen, 2015) and improves on previous estimates in several ways:

i). It incorporates spatial variability within a SLA drawing on spatially distributed forfeited yield, taking account of variable soil factors;

ii). It accounts for temporal variability in productivity, driven primarily by seasonal rainfall variability, to produce climate-adjusted production benefit estimates;

iii). It explicitly accounts for the opportunity cost of resources used in deriving progressive gains in production (marginal costs);

iv). It provides a basis to develop risk-adjusted estimates of likely economic benefits from targeting optimal management of known soil constraints.

The method is only suitable to develop order of magnitude estimates at broad area level, hence the presentation of results at SA2 level. Availability of data constrains the application of more advanced econometric techniques. Although the magnitude of benefits will vary significantly based on the severity of the constraint, management practices, quality of grain and a host of other factors within and beyond farmers' control, these estimates provide useful guidance in determining appropriate scale of interventions.

Results

Magnitudes of yield gaps due to sodicity, acidity, salinity and compaction

The resulting maps of yield gaps due to sodicity, acidity, salinity and compaction are shown in Figure 7. Yield gaps were generally largest for sodicity across much of Australia's cropping land, with average gaps of 0.2–0.4 t ha^{-1} per annum estimated across a number or regions. Yield gaps due to acidity were largest in the high-rainfall regions of Western Australia, Victoria and New South Wales, where they were generally estimated to be around 0.1–0.2 t ha^{-1} per annum on average. Yield gaps due to salinity and compaction were considerably smaller, and predominantly estimated to be less than 0.1 t ha^{-1} per annum as SA2 averages. As averages across Australia's cropping land, the yield gaps due to sodicity, acidity, salinity and compaction were 0.13, 0.04, 0.02 and 0.02 t ha^{-1} per annum, respectively.

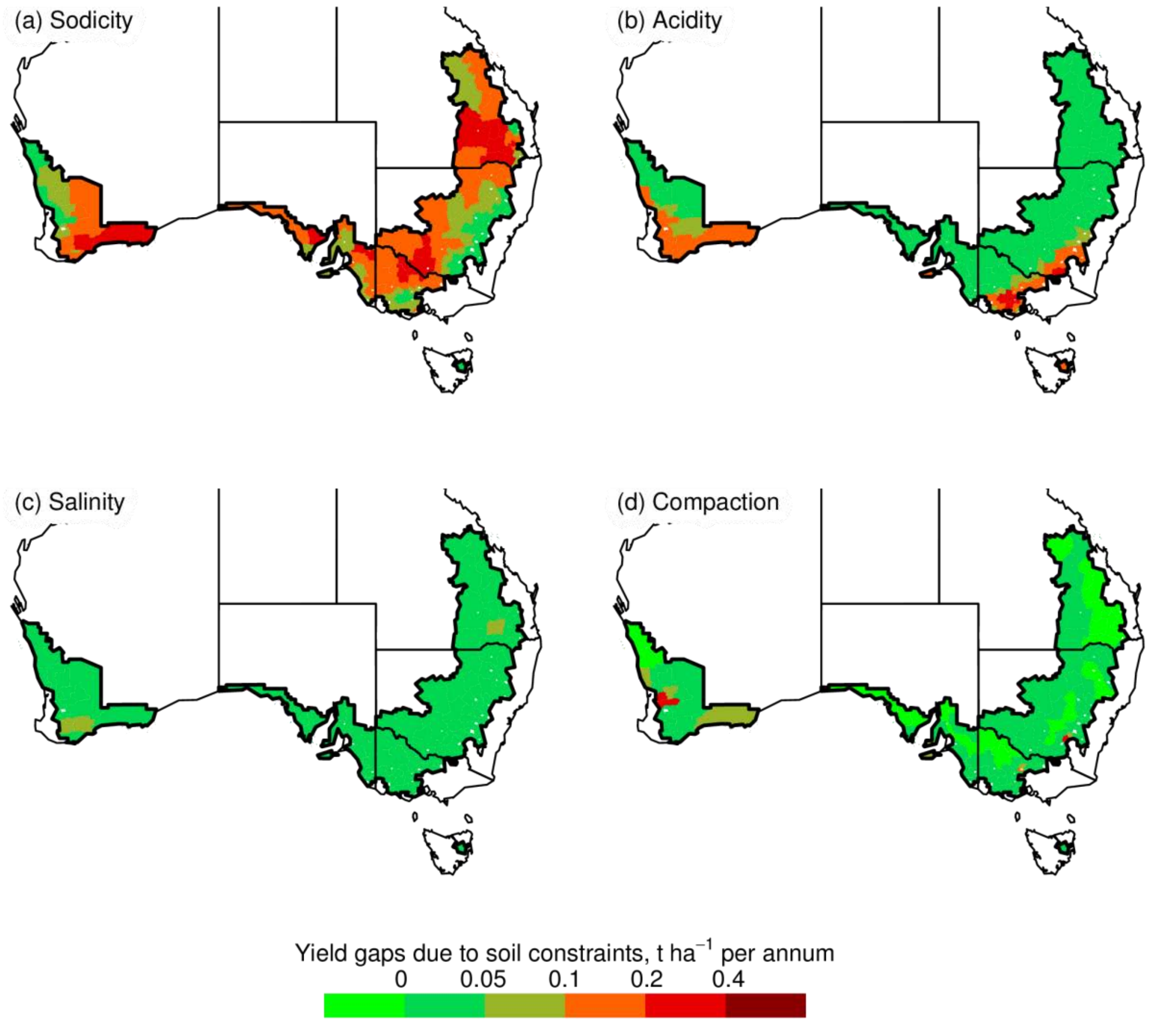


Figure 7 Yield gaps due to (a) sodicity, (b) acidity, (c) salinity and (d) compaction at SA2 level

Areas affected by sodicity, acidity, salinity and compaction

The areas predicted to be affected by each of the soil constraints are shown in Figure 8. For sodicity, regions of Queensland showed more than 90% of cropping land affected, though for much of Australia, in the regions that showed large yield gaps, 75–90% of the cropping land was affected. In the regions affected by acidity (those noted in the previous section), upwards of 90% of the land was predicted to be affected. Salinity was predicted to be affecting yield in the south of Western Australia and the south of Queensland, although Figure 7c suggested that the magnitudes of the yield gaps due to salinity in these regions were not as large as those for sodicity.

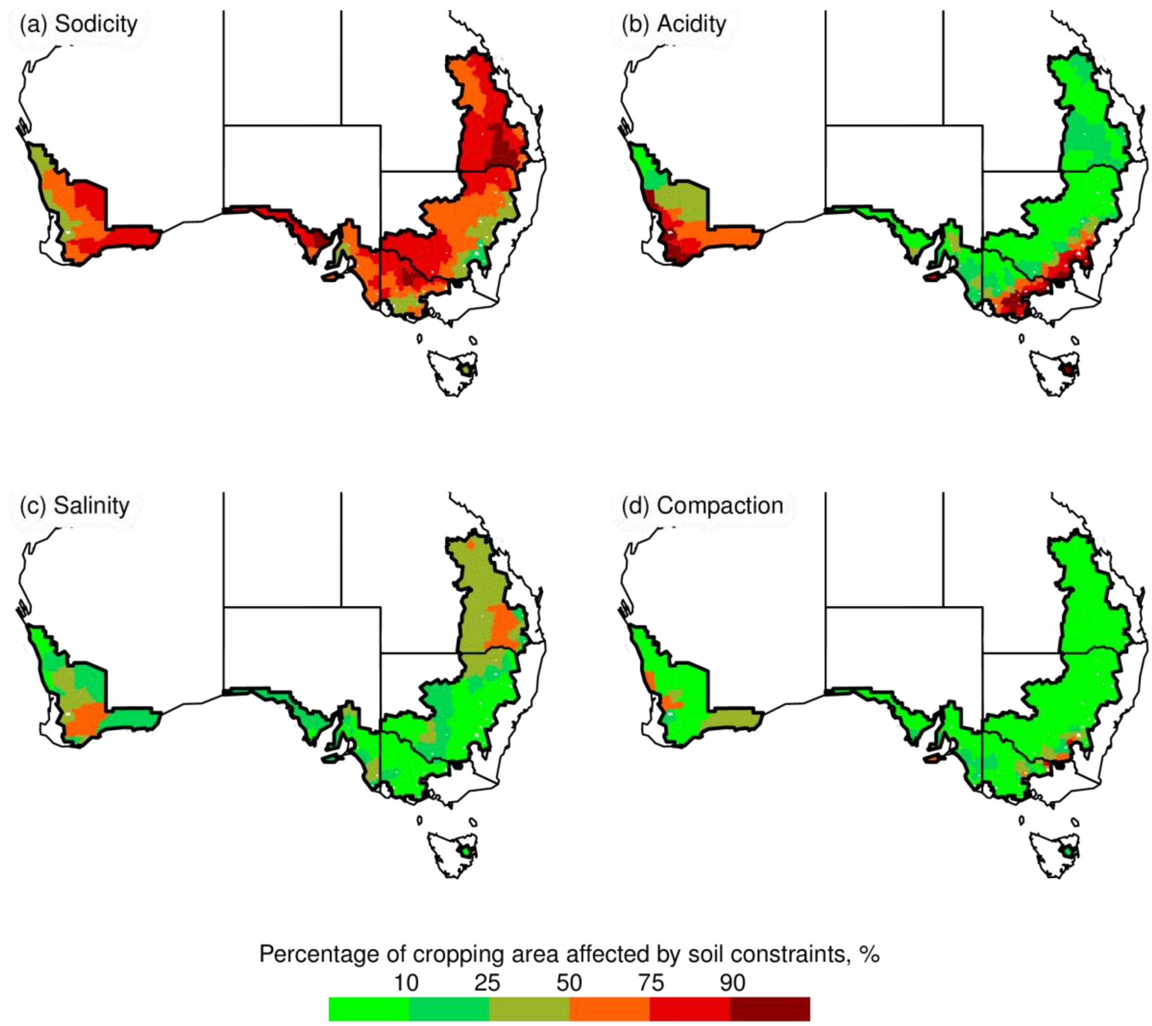


Figure 8 Areas of land, as a percentage of the cropping land, affected by (a) sodicity, (b) acidity, (c) salinity and (d) compaction at SA2 level

Economic impact of soil constraints

Figure 9 shows the gross value of the estimated yield gaps, and Figure 10 the net value, after taking into account indicative treatment costs for sodicity (gypsum) and acidity (lime), both displayed on a per-hectare-of-cropping basis. The grey regions of Figure 10 show where the average yield gap for an SA2 was less than 0.05 t ha^{-1} per annum (no economic analysis carried out) or the predicted net value of ameliorating the soil constraint was less than 10 \$ ha^{-1} per annum (minimal economic benefit). The predicted potential benefits of applying gypsum to ameliorate soil sodicity covered the largest area, with predicted benefits of 20–60 \$ ha^{-1} per annum across much of Australia's cropping land. For acidity, much of Australia fell into the minimal economic benefit category, although there are concentrated parts of Western Australia, New South Wales and Victoria for which potential benefits of 20–60 \$ ha^{-1} per annum have been predicted.

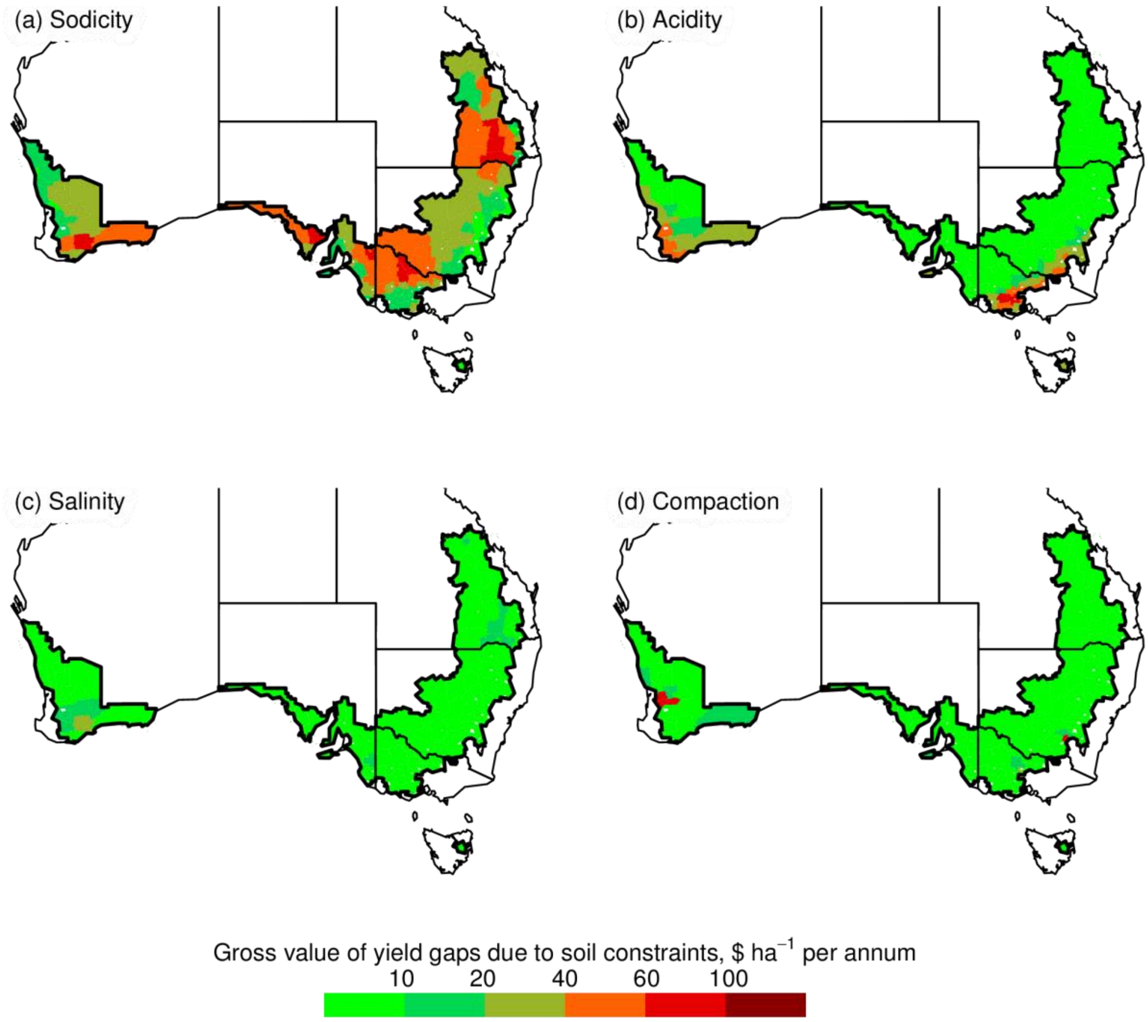


Figure 9 Gross economic value of yield gaps due to (a) sodicity, (b) acidity, (c) salinity and (d) compaction at SA2 level

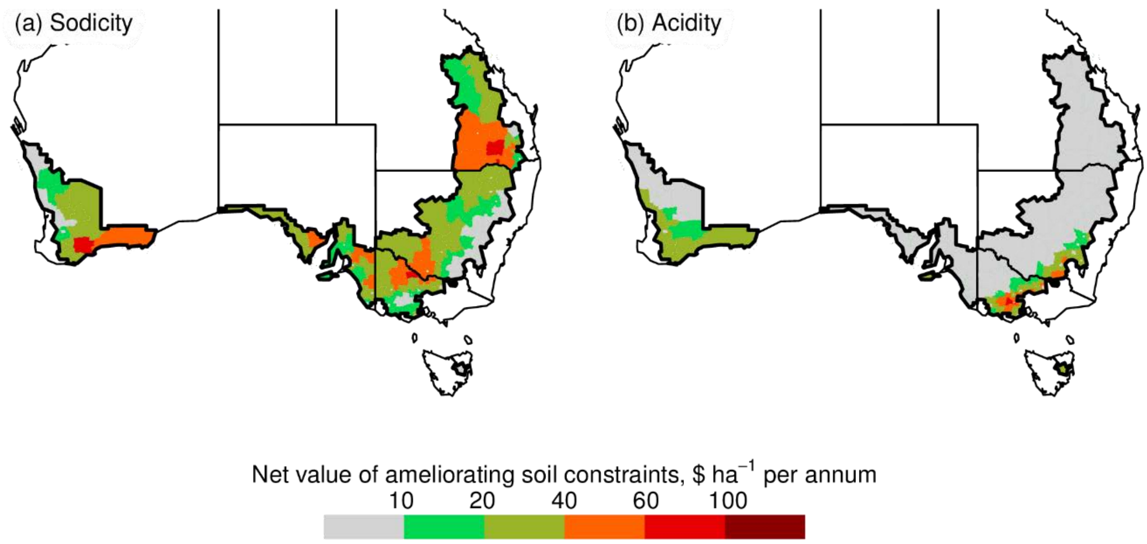


Figure 10 Net economic value of yield gaps due to (a) sodicity and (b) acidity at SA2 level

Broad-scale comparisons

Although the aim of the project was to provide information at the SA2 level, it can be informative to also summarize at broader scales, by state and nationally (Table 6). Results highlight the importance of sodicity nationally and acidity in Western Australia, Victoria and New South Wales. It was estimated that 68% of Australia's grain cropping land is affected by sodicity, 35% by acidity, 24% by salinity and 10% by compaction. In Queensland, sodicity was estimated to be affecting 80% of the cropping land, and the average yield gap due to sodicity was 0.17 t ha^{-1} per annum. In Western Australia, the percentage affected by acidity was 55%, with an average yield gap of 0.07 t ha^{-1} per annum, though we note that this average yield gap includes the areas of non-acidic soil; within just the affected areas, the average yield gap will be larger. In Victoria and New South Wales, 22% of cropping land was impacted by acidity, and the average yield gap was 0.03 t ha^{-1} per annum.

Table 6 Yield gap information, summarized by state and nationally. pa: per annum

State	Area of cropping, kha	Soil constraint, *c*	Average yield gap due to *c*, t ha^{-1} pa	Area affected by *c*, kha	Gross value of yield gap due to *c*, M$ pa	Net value of ameliorating *c*, M$ pa
NSW	8064	Sodicity	0.11	5246	222.2	191.7
		Acidity	0.03	1590	58.6	50.7
		Salinity	0.01	1074	17.6	NA
		Compaction	0.01	586	13.9	NA
Qld	3278	Sodicity	0.17	2631	147.2	126.9
		Acidity	0.01	367	7.5	6.5
		Salinity	0.03	1492	21.7	NA
		Compaction	< 0.01	23	0.7	NA
SA	4964	Sodicity	0.15	3532	195.6	168.7
		Acidity	0.01	504	10.1	8.7
		Salinity	0.01	722	12.8	NA
		Compaction	< 0.01	274	6.1	NA
Tas	108	Sodicity	0.03	24	0.7	0.6
		Acidity	0.18	102	5.1	4.4
		Salinity	< 0.01	1	0.1	NA
		Compaction	0.01	12	0.2	NA
Vic	4614	Sodicity	0.19	3618	223.3	192.6
		Acidity	0.04	1204	48.1	41.6
		Salinity	< 0.01	139	5.7	NA
		Compaction	0.01	386	6.9	NA
WA	17326	Sodicity	0.12	10946	546.7	471.5
		Acidity	0.07	9576	310.8	268.6
		Salinity	0.03	5612	124.4	NA
		Compaction	0.03	2471	129.4	NA
National summary						
	38354	Sodicity	0.13	25997	1335.7	1152.1
		Acidity	0.04	13342	440.3	380.6
		Salinity	0.02	9040	182.3	NA
		Compaction	0.02	3752	157.2	NA

Validation

We investigated the performance of our yield gap estimates using data from four sources: the GRDC's National Paddock Survey (NPS) project, the Queensland section of the National Soil Carbon Research Program (SCaRP), the New South Wales Office of Environment and Heritage's 2008 Monitoring, Evaluation and Reporting (OEH-MER) program, and soil testing data collected by Nutrient Advantage® Laboratory Services of Incitec Pivot Limited (referred to from here as Incitec Pivot). We split this validation exercise into two components. First, we used the soil data from the four datasets in turn to confirm whether the regions that we identified as being affected by each soil constraint (see Figure 8) were indeed subjected to the soil constraint. Second, to provide an indication of whether our predicted yield gaps (see Figure 7) were reasonable, we used just the NPS dataset, which includes both soil data and corresponding yield monitor data.

Validation of areas affected by soil constraints

In the first component of our validation exercise, we used the soil data from all four datasets to provide an indication of whether our predicted areas affected by soil constraints were reasonable. With each dataset in turn, we considered only SA2s containing data from at least four separate locations. For these SA2s, we considered the data from each location in turn, and assigned that location as 'affected' or 'unaffected' by a soil constraint depending on whether the critical values for that soil constraint (see Table 4) were exceeded for any of the three soil depths. Our validation datum for that SA2 was then the proportion of locations in that SA2 classified as 'affected', and this value was compared with our predicted proportion of that SA2's cropping land affected by the soil constraint (the values mapped in Figure 8). For the Incitec Pivot dataset, exact spatial coordinates were unavailable and data were attributed to postal areas instead; in this case, we therefore compared predicted and observed proportions of each postal area affected. Furthermore, the Incitec Pivot dataset did not contain information to link topsoil samples with subsoil samples from the same soil profile, and it was therefore assumed that all samples were from separate profiles, with each measurement assigned to one of our three harmonized soil depths (A: 0–10 cm, B: 10–50 cm, C: 50–200 cm) based on the sampling interval midpoints.

We also note here that the validation data do not represent exactly the same concept as our predicted areas affected, Figure 8, which not only represent a probability of the critical values being exceeded, but also take into account whether or not the soil constraint should be expected to have an impact on yield for a given climate and soil texture (according to the fitted model). Nevertheless, we use the exercise to give some idea of general agreement of areas where the soil constraints could potentially be a problem.

Figure 11 shows the predicted proportion of an SA2 affected by each soil constraint (y-axes) plotted against the observed proportion affected (based on each validation dataset separately) for sodicity, acidity and salinity. The figures show some general agreement, in that the areas with larger observed proportions affected also had larger predicted proportions affected. This is also demonstrated by the generally positive values of Lin's concordance correlation coefficient (CCC; Figure 11), a measure of the agreement between the observed and predicted values. For soil sodicity, our predicted proportions tended to be larger than the observed proportions. One possible reason for this (in the case of the NPS validation data) is that the NSSC data used to build the model on which the predicted area affected was based have higher ESP values than the NPS data: 14%, 36% and 46% of the NPS data were classed as sodic (ESP > 6%) for the harmonized depths A, B and C, respectively, while the corresponding percentages in the NSSC dataset were 23%, 58% and 76%, respectively. For soil acidity, validation against the OEH-MEH data suggested that (in New South Wales) model predictions of the area affected were too small. However, as noted before, this could be due to the predictions not only taking into account the soil acidity, but also the likelihood of it having an impact on yield for a given climate and soil texture. Validation against the SCaRP data suggested that the predicted area affected by salinity in Queensland was too large. This could be due to differences between the validation and calibration datasets, the former of which were collected to a depth of 30 cm and did not include Cl data, the latter of which also included data below 50 cm and measurements of Cl. Previous work has highlighted the issue of subsoil Cl in Queensland (e.g. Dang et al., 2010), which would not be detected by the validation data to 30 cm, and this could be the reason that predicted areas of salinity were larger than observed areas in the SCaRP dataset in particular.

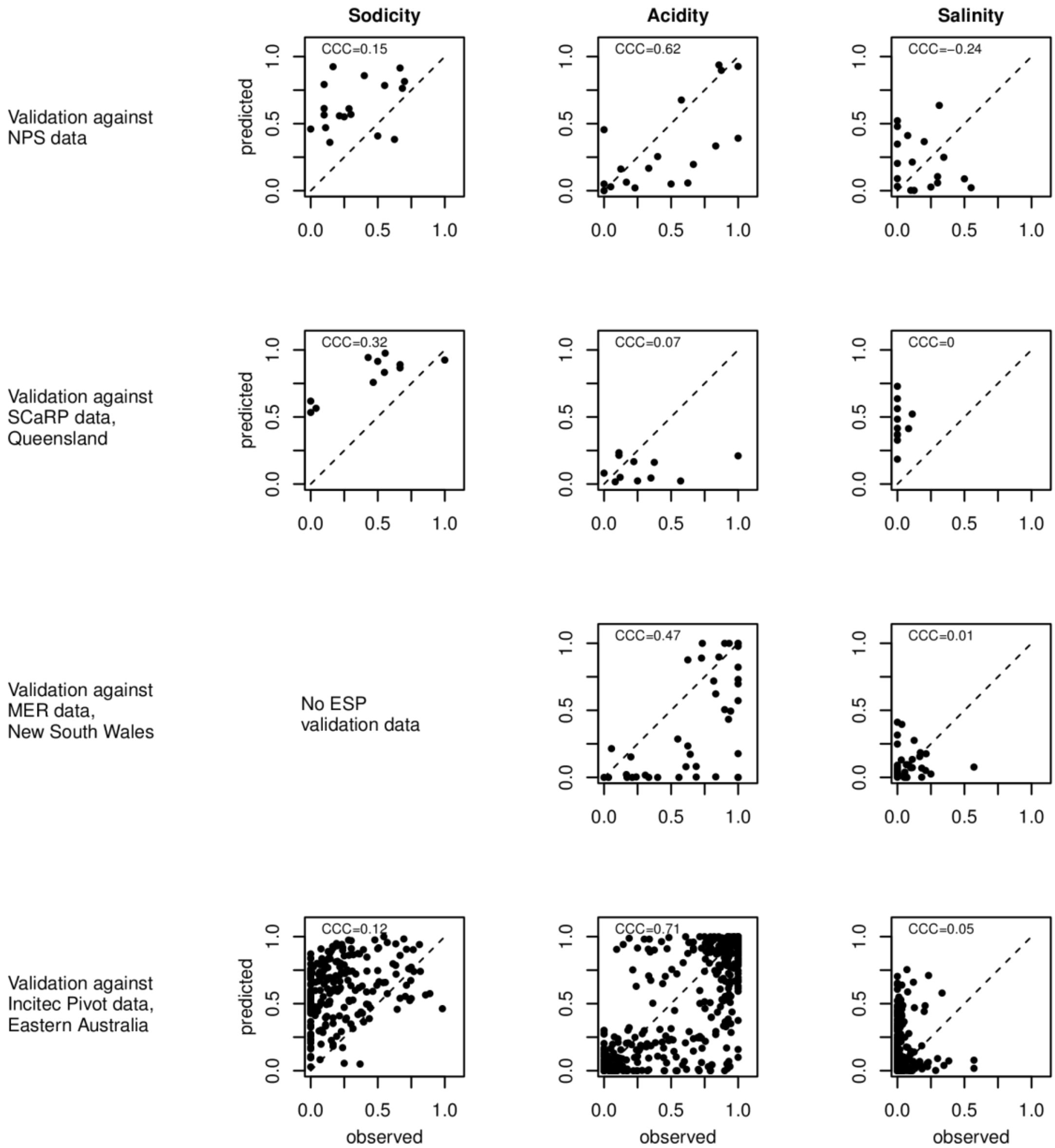


Figure 11 The predicted (y-axes) against observed (x-axes) proportions of each SA2 (with sufficient validation data) affected by soil sodicity (left), acidity (centre) and salinity (right) for the NPS (top row), SCaRP (second row), OEH-MEH (third row) and Incitec Pivot (bottom row) validation datasets

Validation of yield gaps due to soil constraints

In the second component of our validation exercise, we used the NPS soil and yield data to investigate the performance of our yield gap estimates. The NPS dataset consists of data from 156 paddocks across Australia, with two zones (labelled A and B) defined as separate transects within each paddock. Five soil cores were collected from evenly spaced locations along each transect and each core split into four sampling depths (most commonly 0–10 cm, 10–40 cm, 40–70 cm and 70–100 cm); the five samples were bulked for each depth separately and sent to laboratories for analysis of a number of soil properties, including ESP, pH, EC and Cl. The dataset also contains yield monitor data, which were processed in the vicinity of the soil sample locations to estimate, for each zone of each paddock, an average yield over the five 30-m blocks centred on the five sampling locations. Differences in yield between the two zones of a paddock could not be attributable to climate, and management differences between the two zones were generally limited (crop varieties and sowing dates were the same for both zones of a paddock in all cases). Therefore yield differences between zones must be attributable to other factors, most notably soil constraints.

Seventy-nine of the paddocks involved in the NPS presented yield monitor data for a wheat crop in 2015 (there were at most 17 paddocks that presented wheat data for any other single year). Of these, the yield monitor data from 45 paddocks gave sufficient coverage to give reasonable estimates of yield for the two zones, and it was these data that we used in this component of our validation exercise.

For each paddock in turn, we used the NPS soil data and climate data (extracted from SILO) together with our fitted Cubist models (with parameters presented in Appendix B) to predict the yield for the two zones of the paddock for the year 2015. We then calculated the difference in predicted yield between the two zones, and compared this to the difference in actual yields (calculated from the yield monitor data) between the two zones. , In each case the predicted yield difference is taken to be the larger of the two predictions minus the smaller, with the validation data calculated to represent the same difference. Thus, if zone A was predicted to have larger yield than zone B for a paddock, the prediction and validation data would both represent yield for zone A minus yield for zone B, and *vice versa*. We carried this validation out for the sodicity, acidity and salinity models, and in each case we only present results from paddocks where soil constraints for at least one of the two zones exceeded the critical values (Table 4).

Table 7 gives the mean predicted and mean observed yield differences, the former based on the models fitted to represent the effects of soil sodicity, acidity and salinity. By definition, the mean predicted yield differences are positive; it is at least somewhat encouraging that the means of the observed differences for these paddocks are also positive, albeit based on a very limited sample size (due to the small number of paddocks where soil constraints were potentially present). Table 8 also gives the number of paddocks for which the predicted yield difference between the two zones was of the correct sign (i.e. if the observed yield difference was positive, then so too was the predicted yield difference, and *vice versa*). In all three cases, the signs of the predicted differences were at least as good as random assignment; in the case of soil acidity, which presented the most validation data, 15 of 26 yield differences were correctly classified as positive or negative. However, with such a limited number of validation sites available where at least one of the zones was affected by the soil constraints, mean observed yield differences in particular had large standard errors (shown in parentheses in Table 7), and robust conclusions cannot be drawn.

Table 7 Validation results comparing predicted and observed yield differences for paddocks where at least one zone was potentially affected by the soil constraint (standard errors in parentheses)

Soil constraint	Number of paddocks with at least one affected zone	Mean predicted yield difference	Mean observed yield difference	Number of correctly classified yield differences
Sodicity	8	0.08 (0.02)	0.08 (0.16)	4
Acidity	26	0.03 (0.01)	0.17 (0.12)	15
Salinity	1	0.02	0.62	1

Sensitivity of yield gap estimates to critical values

The critical values, summarized in Table 4, represent values of each soil constraint in each soil depth beyond which some detrimental effect of the soil constraint on yield can be modelled; for instance, in the topsoil (0–10 cm), a pH below 6.0 can (in the model) result in yield penalties, while in the 10–50 cm soil depth, an EC of more than 0.7 dS m^{-1} can induce penalties. We investigated the sensitivity of our yield gap estimates to these critical values, first by considering critical values that represent a greater tolerance of wheat to the soil constraints (i.e. a smaller pH critical value for the effects of acidity and larger critical values for all other constraints), and second by considering critical values representing greater sensitivity (i.e. the reverse of the above). The ‘tolerant’ and ‘sensitive’ critical values are summarized in Table 8. The critical value of ESP of 6% used in the main study for effects of sodicity corresponded to the classification of a ‘sodic’ soil in Northcote and Skene (1972); their classification of a ‘strongly sodic’ soil was based on an ESP of over 15%, and we therefore consider this value for our ‘tolerant’ critical value of ESP. This value is also the value of ESP above which sodicity effects are represented in the subsoil constraints function developed in Hochmann et al. (2007) for APSIM. Other ‘tolerant’ and ‘sensitive’ critical values were selected judiciously based on the values used in the main study. The critical values for the effects of alkalinity were left at their original values.

Table 8 Summary of (a) ‘tolerant and (b) ‘sensitive’ critical values for all soil constraints and for the three depths, *d*; A: 0–10 cm, B: 10–50 cm, C: 50–200 cm.

d	$ESPCrit_{Sdcty}[d]$	$pHCrit_{Acdty}[d]$	$pHCrit_{Alkty}[d]$	$ECCrit_{Slnty}[d]$	$ClCrit_{Slnty}[d]$	$BDCrit_{Cmpctn}[d]$
	(a) Critical values with wheat more tolerant to soil constraints					
A	15.0	5.5	7.4	0.5	400	1.8
B	15.0	4.5	7.4	1.0	800	1.8
C	15.0	4.5	7.4	1.0	800	1.8
	(b) Critical values with wheat more sensitive to soil constraints					
A	3.0	6.3	7.4	0.2	200	1.4
B	3.0	5.0	7.4	0.5	400	1.4
C	3.0	5.0	7.4	0.5	400	1.4

Results from this sensitivity analysis are presented as maps of yield gaps in Figures C1 and C2 and are summarized at broad scale in Tables C1 and C2 in Appendix C for the ‘tolerant’ and ‘sensitive’ critical values, respectively. Nationally, the impact of sodicity remained the largest, with average yield gaps varying from 0.07 t ha^{-1} per annum (tolerant) to 0.19 t ha^{-1} per annum (sensitive), compared to the value of 0.13 t ha^{-1} per annum based on the critical values used in the main study. The total area affected by sodicity varied between 21 million ha (tolerant) and 31 million ha (sensitive) of the 38 million ha of potential cropping land. For acidity, average yield gaps ranged from 0.03 t ha^{-1} per annum (tolerant) to 0.07 t ha^{-1} per annum (sensitive), with little variation in the total area affected (13 to 16 million ha). Average yield gaps for salinity were not largely affected by changing the critical values of EC and Cl, though the areas affected did vary from 5 to 11 million ha. For compaction, the tolerant critical value of BD of 1.8 g cm^{-3} led to compaction not being included in the model, and therefore not contributing to yield gaps, while the sensitive critical value of 1.4 g cm^{-3} resulted in an average yield gap due to compaction of 0.05 t ha^{-1} per annum.

An alternative approach based on APSIM

APSIM modelling

We investigated an alternative approach for mapping yield gaps due to soil constraints based on a mechanistic crop yield model (APSIM) rather than an empirically-built one. In order for the APSIM-based approach to give predictions of the same quantity as the empirical approach, we attempted to set up APSIM under some locally general management conditions, so that it could as closely as possible predict the spatial distribution of actual crop yields around Australia. Subsequently, we followed an analogous approach to that used in the empirical modelling approach: first APSIM was run using actual soil constraints data from each of the NSSC profile locations to give y_{ac}, and second it was run using the same input data except for optimized values of the soil constraint of interest to give y_{oc}. Again, the difference between these two model runs (i.e. Equation 2) gave us the yield gap due to soil constraint c at data locations, and these yield gaps due to soil constraints were interpolated and aggregated to give predictions at the spatial support of SA2. Further details on this APSIM approach are presented in Appendix D.

We note here that this approach based on APSIM has been implemented for comparative purposes only, and that we base our conclusions from the project on the results from the empirical approach. Some aspects of the approach that could influence results should also be noted. The models representing the effects of soil constraints in APSIM have not been tested in a wide range of situations. Indeed, the effects of sodicity and salinity are represented by a function that alters the APSIM parameter KL, the crop's potential water-extraction rate from a soil layer (Hochmann et al., 2007). This function was originally built in to represent the effects of subsoil constraints in particular, with data from 33 paddocks on Vertosols in the Northern Grains Region (New South Wales and Queensland) used to fit appropriate relationships between the subsoil constraints and the APSIM parameter KL. Our use of this function to estimate yield gaps due to sodicity and salinity across the entirety of Australia's wheat-cropping land represents a significant extrapolation. Similarly, the function of pH that we used to represent the effects of soil acidity (through its impact on the root exploration factor, the APSIM parameter XF) was built by combining data from different sources and its applicability for use in the manner we have across Australia is far from certain. Nevertheless, results from this approach demonstrate the possibility of alternate methods to estimate the same quantity, and regardless of any numerical differences could provide some useful insight into spatial distributions of yield gaps due to soil constraints and their relative importance.

Figure 12 shows the yield gaps due soil sodicity, acidity and salinity based on the APSIM approach. The average yield gaps due to these soil constraints across Australia's cropping land were 0.08, 0.09 and 0.06 t ha^{-1} per annum for sodicity, acidity and salinity, respectively. The figure for sodicity in particular was considerably smaller than the corresponding figure of 0.13 t ha^{-1} per annum based on the empirical approach. One reason for this is the function used by APSIM to represent the effects of soil sodicity through its impact on the parameter KL. This function represents effects of soil sodicity on yield only when the ESP exceeds 15% (Hochmann et al., 2007), compared with the critical value of 6% applied in our empirical approach. Furthermore, the function was developed for representing the effects of subsoil constraints, and it may not be able to represent effectively the impact of sodicity in the topsoil, where the ESP is typically smaller than in the subsoil. The figure does, however, align well with the national average yield gap due to sodicity of 0.07 t ha^{-1} obtained in the sensitivity analysis when a critical value of 15% was used to define effects of sodicity (Figure C1 and Table C1 in Appendix C).

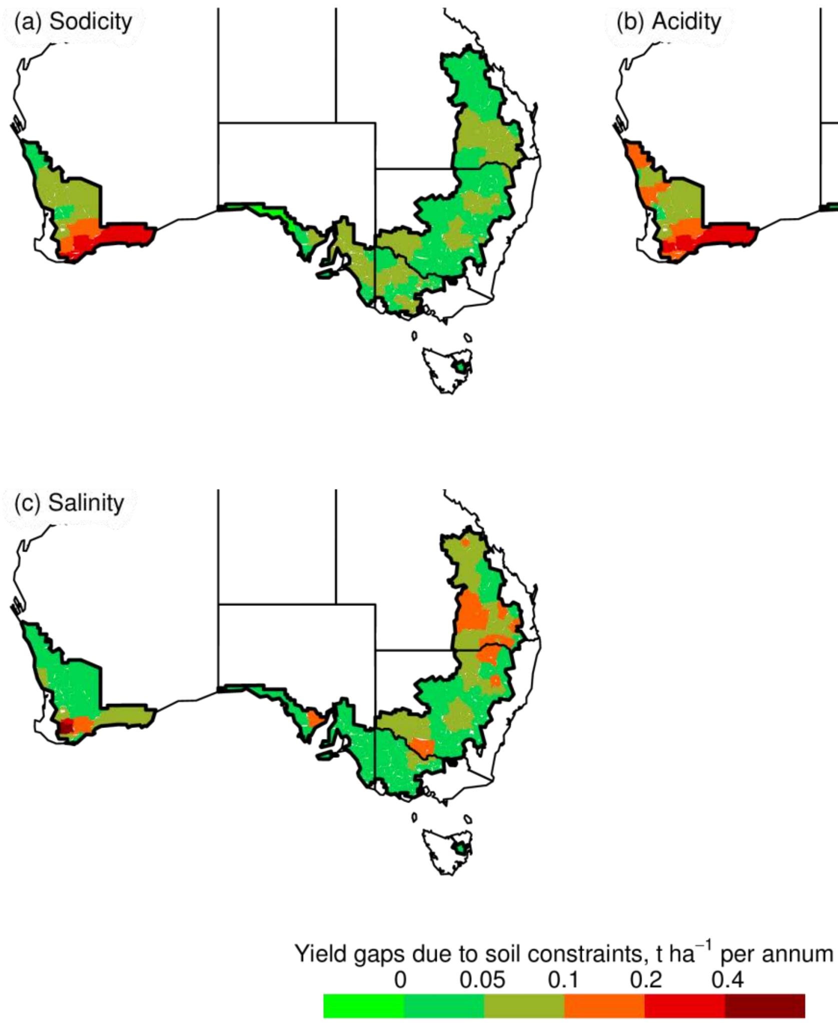


Figure 12 Yield gaps due to (a) sodicity, (b) acidity and (c) salinity at SA2 level estimated by the APSIM approach

Combining predictions from the two approaches

In addition to APSIM providing an alternative approach, it is possible to combine results from the Cubist (empirical) and APSIM modelling approaches. To demonstrate, consider the yield gap estimates at soil data locations produced by the Cubist and APSIM methods, denoted here by $Y_{gc}^{(C)}\left(\mathbf{x}_\mathrm{S}^{(C)}\right)$ and $Y_{gc}^{(A)}\left(\mathbf{x}_\mathrm{S}^{(A)}\right)$, respectively. Due to different levels of data requirements for the two approaches, not all soil data locations produced yield gap estimates by both approaches (i.e. , $\mathbf{x}_\mathrm{S}^{(C)}$ and $\mathbf{x}_\mathrm{S}^{(A)}$ are not identical sets of locations). (For instance, for APSIM modelling we imposed the requirement that soil data were available to a depth greater than 50 cm, whereas the iterative approach adopted for fitting Cubist models in the empirical approach, see Appendix A, meant that models could be fitted without such data.) Therefore, an interpolation procedure that accounts for the correlation between the estimates from the two approaches, $Y_{gc}^{(C)}\left(\mathbf{x}_\mathrm{S}^{(C)}\right)$ and $Y_{gc}^{(A)}\left(\mathbf{x}_\mathrm{S}^{(A)}\right)$, should prove beneficial, particularly in areas where estimates were only able to be made based on one of the two approaches. Cokriging, with a correlation structure modelled through a linear model of co-regionalisation (LMCR; Webster and Oliver, 2007), provides such an interpolation procedure, and we adopt this approach here. This results in two sets of predictions for the interpolation grid locations, $\mathbf{x}_\mathrm{k}$: $Y_{gc}^{(C)}(\mathbf{x}_\mathrm{k})$ denoting the yield gap predictions based on the Cubist approach and $Y_{gc}^{(A)}(\mathbf{x}_\mathrm{k})$ denoting those based on APSIM. Note the benefit of these predictions over those presented in Figures 7 and 12, respectively, is that now the correlation between $Y_{gc}^{(C)}(\mathbf{x}_\mathrm{S}^{(C)})$ and $Y_{gc}^{(A)}(\mathbf{x}_\mathrm{S}^{(A)})$ is utilised in the interpolation procedure. To produce a single set of yield gap predictions, one could consider a weighted average of $Y_{gc}^{(C)}(\mathbf{x}_\mathrm{k})$ and $Y_{gc}^{(A)}(\mathbf{x}_\mathrm{k})$. Figure 13 presents the SA2-aggregated maps based on such weighted averages, using equal weights for the predictions produced using the two methods, $Y_{gc}^{(C)}(\mathbf{x}_\mathrm{k})$ and $Y_{gc}^{(A)}(\mathbf{x}_\mathrm{k})$.

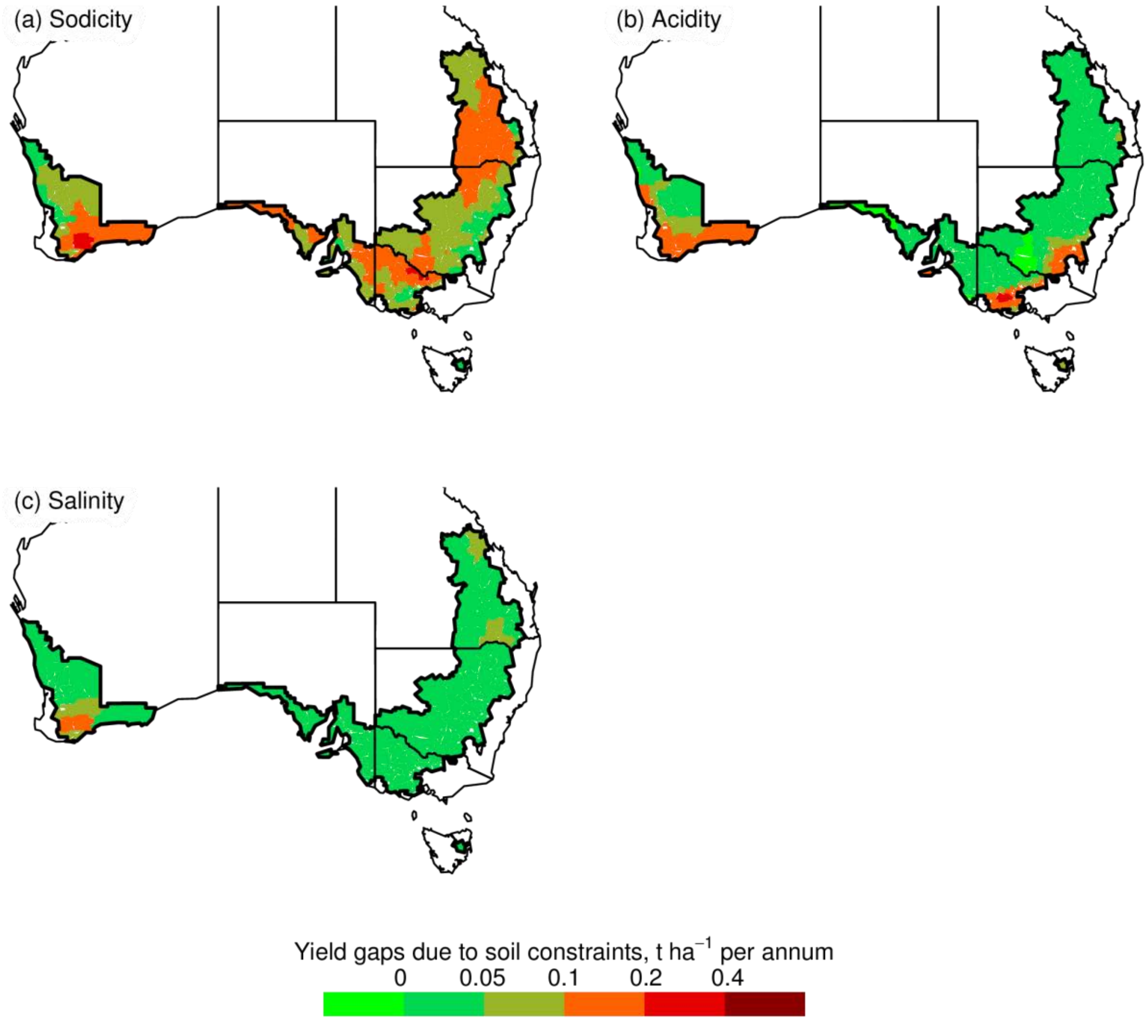


Figure 13 Yield gaps due to (a) sodicity, (b) acidity and (c) salinity at SA2 level estimated by cokriging yield gaps from the Cubist (empirical) and APSIM approaches, and then averaging the two sets of predictions

Discussion of Results

Yield gaps due to soil constraints

The results we have presented suggest that of the four soil constraints considered, sodicity has the largest effect on wheat yields across Australia, with acidity also producing large yield gaps in the high-rainfall areas of Western Australia, Victoria and New South Wales. Hajkowicz and Young (2005) also found these two constraints to have larger impacts on wheat yield when compared with those of soil salinity. In a previous GRDC project, van Gool (2016) looked at the occurrence of a number of soil constraints across Western Australia, and in particular presented maps of the occurrence of topsoil and subsoil acidity, surface structural decline (often soils with variable levels of sodicity), salinity and subsurface compaction, which appear to show reasonable visual agreement with those presented here for the areas affected by soil constraints (Figure 8).

In previous work, estimates have been derived for the total yield gap due to genetic and management factors (Hochmann et al., 2012). On average our estimated yield gaps due to sodicity and acidity represent 9 and 4% of these total yield gaps, respectively.

We have not yet implemented yield gaps due to all soil constraints; this would be possible within the current methodology, by simultaneously fitting a model representing the effects of all soil constraints. However, for consideration of the potential benefits of individual treatments in the current work, we investigated each soil constraint in isolation, and the combined effect of all soil constraints provides scope for further work.

Economic value of addressing soil constraints

Hajkowicz and Young (2005) reported potential annual profit increases of $1000 million, $1600 million and $200 million for the costless removal across the Australian wheat cropping belt of soil sodicity, acidity and salinity, respectively. The corresponding values estimated in the current work (to the nearest $100 million) are $1300 million, $400 million and $200 million, respectively, the most notable difference being the value of yield gap due to acidity. The difference between these estimates would be due in part to the extent of remediation already undertaken by land managers, in particular for soil acidity, as well as refinements in the method of assessment.

Soil acidification is a natural process accelerated by agricultural practices. The main cause of soil acidification is inefficient use of nitrogen, followed by the export of alkalinity in produce. Ammonium-based fertilisers are major contributors to soil acidification. The treatment of acidity is reasonably straightforward, liming being the most economical method of amelioration. The amount of lime required will, however, depend on the soil pH profile, lime quality, soil type, farming system and rainfall.

In recent years, an improved understanding of cropping systems, including plant physiology and determinants of soil fertility, has enabled farmers to better target fertiliser and crop chemicals, especially nitrogen and soil ameliorants such as lime and gypsum for better effect. The use of liming material to control acidity is now widespread and the number of agricultural businesses applying lime, dolomite and other types of soil ameliorants in Australia increased 19 per cent between 2013–2014 and 2014–15. In 2014–15, an estimated 3 million tonnes of lime were applied to 2.3 million hectares of land across Australia, with the majority used in Western Australia (ABS, 2016).

We must note that our estimates of the economic impacts of soil constraints are based on the entirety of Australia's 'potential cropping land' (that classified as dryland or irrigated cropping in the ACLUMP map). Across Australia this covers 38 million ha, whereas on average only 13 million ha is cropped with wheat in any given season. Therefore in reality our estimates provide an overestimate of the likely increase in profits if the soil constraints were to be ameliorated entirely through the treatments considered here. In considering future benefits of amelioration, this ‘optimistic’ potential area may prove useful in any given region as the wheat crop often moves around available land following rotations and seasonal conditions. Once ameliorated, the land is available for production, offering potential benefits.

The net estimates of the potentially forgone value of agricultural production due to principal soil constraints, soil sodicity and soil acidity, produced in this study need to be considered in the broader context of Australian farm financial performance. For instance, the estimated gross value of production for the grains industry in 2014–15 was around $14 billion. Wheat accounts for around half of this value as well as the volume of grains production (Ashton et al. 2016).

In evaluating measures to address existing yield gaps that have been identified in this work and many previous assessments, a key step is to ascertain the nature and severity of any barriers that may prevent profit maximising farmers adopting measures that will profitably address these soil constraints potentially contributing to yield gaps. In doing so, the marginal cost approach that we have adopted in this analysis will prove sub-optimal.

Methodology

Lobell (2013) outlined general approaches that might be applied to understand yield heterogeneity at the landscape scale, and our approach is based on one of these. That is, maps of yields derived from remote sensing are compared with ancillary datasets on factors thought to control yields, in our case climate and soil properties. Statistical analyses are then used to evaluate the relative importance of each factor in driving observed yield variations.

A few caveats should be noted for the interpretation of results following this method. The methodology was designed to make the best use of data available to this project in order to produce information about yield gaps at large spatial scales (SA2). However, there are a number of steps in the process, all of which will carry some level of uncertainty. These include uncertainty in the:

- ABS SLA-level yield data
- disaggregated yield predictions
- soil data (resulting from measurement errors and the soil depth harmonization process)
- Cubist model fitted to the data
- estimated yield gaps at the soil data locations (as a result of all the former uncertainties)
- interpolated yield gaps

We have attempted to capture many of these sources, for instance by refitting the Cubist model as a linear mixed model (LMM) that better accounts for parameter uncertainty, and by carrying out the interpolation on a log scale so that the statistical model behind the interpolation effectively represents data uncertainties (i.e. the uncertainties in the yield gaps at the soil data locations) that are larger for data of larger magnitudes. However, we have not accounted for all of these sources, and as such future work could address some of these issues and aim to integrate them all together into the final yield gap estimates. We also note here that although our approach to refit the Cubist model as a LMM accounts for the basic structure of the dataset, we have not implemented the approach to fully account for spatial correlation; this too provides scope for further work.

As the above uncertainties in estimation of yield gaps are carried in to the economic analysis, the decision was taken to eliminate all SA2s with very small yield gaps (<0.05 t ha^{-1} per annum) for the economic analysis. Similarly, the economic analysis relies on cost estimates and prices of output that ignore regional variability, product quality and management factors that would affect the costs and returns. Therefore, the estimates produced need to be taken in the light of these uncertainties and as a guide to decision making at appropriate scales, such as SA2 and above.

It should be noted that this study did not implement the final step of evaluating the net economic benefits of amelioration for soil salinity and soil compaction. This was due to the unavailability of a specific agronomic measure to address these soil constraints. Without such a measure, it is not meaningful to relate the costs and benefits of amelioration to assess economic benefits. Furthermore, soil salinity can result in the permanent loss of productive capacity, for which no management options would be able to return the land to a profitable wheat-growing state. Our analysis would have excluded any data from such land from model fitting; such land might already have a non-cropping classification in the ACLUMP land use map, and if not, our time-series analysis of remotely-sensed vegetation index data will most likely have classified all pixels within the affected area as not actively cropped (due to a lack of evidence of Werker-Jaggard-type growth). This could be a reason for the relatively low cost attributed to soil salinity in this work, and alternative methods would be required to identify and cost this type of land degradation issue.

While the authors acknowledge the availability of estimates for those soil constraints from previous work, it is unwise to develop estimates using methods that cannot be substantiated at the scale of this work. However, the method adopted offers a robust framework to carry out benefit estimates for those constraints, if data supporting specific amelioration methods can be identified. That will be an area for further work, as will be developing methods to apply amelioration at a more local scale using area specific cost estimates.

Presentation of results

In this report, we have presented results as maps at SA2 level of soil constraints and their costs to the Australian grains industry, which should be interpreted with care. The SA2s themselves are of vastly different sizes (those within the cropping region vary in surface area from 1200 to 6 million ha), and furthermore vary considerably in the intensity of cropping land (from less than 1% to 100% cropped). Therefore data presented in such maps may create biases in interpretation. For instance, the importance of an SA2 with large surface area and very sparse cropping could be overemphasized compared to a small SA2 with concentrated cropping activity. One alternative to mapping the net value of ameliorating the soil constraint on a per-hectare-of-cropping basis would be to map total net values instead of per hectare net values; however, in this instance large, intensively-cropped SA2s would show large total values and the map would not shed real light to answer our question, where is amelioration of the soil constraint most important? To overcome this problem, data relating to maps have also been tabulated and made available in interactive Excel spreadsheet form, whereby one can select any particular SA2 of interest and explore the full set of data for that SA2. Furthermore, the impact of the critical values on yield gap estimates can be explored through this interactive table, so that the user can select either the standard (Table 4), tolerant (Table 8a) or sensitive (Table 8b) critical values to apply for each of the soil constraints. Any SA2s for which more than 25% of their potential cropping land was classed as irrigated (according to the ACLUMP land use map) have been flagged, to indicate areas where results could be affected. Although it is difficult to communicate this table in a report, it should provide a valuable resource for GRDC to draw robust insights from this study.

Conclusion

Of the four soil constraints considered in this work—sodicity, acidity, salinity and compaction—sodicity gave the largest magnitude of yield gaps across Australia, with an average yield gap of more than twice that of acidity. Yield gaps due to acidity were more concentrated spatially, in the high-rainfall regions of Western Australia, Victoria and New South Wales. Across the wheat-growing land of Australia, the total potential annual economic benefit of ameliorating these soil constraints was estimated to be around $1.15 billion per annum for sodicity (application of gypsum) and $380 million per annum for acidity (application of lime). We note that these are based on indicative costs only, and are intended to provide information at broad rather than fine spatial scale.

Implications

With an increasing population and greater stresses on food production, it is becoming more important to utilize the land as efficiently as possible. There are potential increases in yield and economic benefits to be earned from investment in strategies to combat soil constraints, and these benefits should contribute towards ensuring greater profits for farmers and better food security for our future.

Recommendations

- Invest in strategies for managing soil sodicity: technologies and strategies for the application of gypsum and other ameliorants
- Invest in strategies for managing soil acidity in regions where deemed to be the problem.
- Speak to farmers in the regions deemed to be most heavily affected by particular soil constraints. Confirm if appropriate management of these constraints would be feasible for them.

The Project Team

UQ: Yash Dang, Thomas Orton, Professor Neal Menzies, Professor Ram Dalal, Associate Professor Peter Kopittke, Thilak Mallawaarachchi

CSIRO: Zvi Hochman, Ross Searle, Neil McKenzie, Peter Wilson, Neil Macleod, Roger Lawes

DSITI: Matthew Pringle

DNRM: Andrew Biggs

UA: Pichu Rengasamy

NSW Environment: Jonathan Gray

DPIRD WA: Paul Galloway

Equii WA: Wayne Pluske

DEPI Victoria: Mark Imhof

Cropfacts: Harms van Rees

Incitec Pivot: Charlie Walker, Paul Kennelly

GRDC: Jan Edwards, Gordon Cumming

Appendix A. Cubist modelling details

Handling missing data

Missing covariate data were dealt with by first removing rows of the dataset with any missing values and then fitting the model. Subsequently, columns (variables) were removed from the complete dataset that were not required by the fitted Cubist model. These two steps were iterated until all variables presented to the Cubist software were included in the fitted Cubist model.

We first defined the full dataset of *n* entries of *p* covariates contained in the $n \times p^{(0)}$ matrix, $\mathbf{W}^{(0)}$. Then, for $i = 1,2, \ldots$, the following steps were iterated:

(i) Define the $n^{(i)} \times p^{(i)}$ matrix $\mathbf{X}^{(i)}$ by removing all rows of $\mathbf{W}^{(i-1)}$ containing any missing data

(ii) Fit a Cubist model, $\mathbf{M}^{(i)}$, with covariate data $\mathbf{X}^{(i)}$

(iii) If all variables presented to the model (i.e. those contained in $\mathbf{X}^{(i)}$) are used (either as split variables or as variables in the multiple regression models within rules) in model $\mathbf{M}^{(i)}$, stop; otherwise, define the $n \times p^{(i+1)}$ ($p^{(i+1)} \le p^{(i)}$) matrix $\mathbf{W}^{(i+1)}$ by removing from $\mathbf{W}^{(0)}$ any variables that were not used in $\mathbf{M}^{(i)}$, increment i, and return to step (i)

Note that for $i \ge 1$, we have $n^{(i+1)} \ge n^{(i)}$, so that the model is fitted with more entries as more redundant variables get removed from the set of variables presented to Cubist.

Accounting for correlation structure

Assume that a Cubist model has been fitted to model the yield, y_{it}, for location i in growing season t. Also assume that the model has R rules, and write I_{jit} for the indicator variable that takes the value 1 if rule j defines a subset of the data containing datum y_{it}, and the value 0 otherwise. Define $R_{it} = \sum_{j=1}^{R} I_{jit}$, the number of rules defining subsets containing y_{it} (it is possible for any datum to fall under more than one rule). Further, let the multiple regression model that applies under the conditions of rule j be based on the values of the p_j covariates w_{jkit}, $k = 1, \ldots, p_j$ (which include a constant term for $k = 1$ with each j). Then a LMM can be written for the data:

$$y_{it} = \sum_{j=1}^{R} \sum_{k=1}^{p_j} \frac{I_{jit}}{R_{it}} w_{jkit} \beta_{jk} + b_i + c_t + \varepsilon_{it}$$

where the summation defines the fixed-effects part of the model (i.e. if β_{jk} were taken to be the parameter values fitted by Cubist, this expression would give exactly the Cubist predictions at the data locations). The remaining terms give the random effects and residual error: b_i is a location-specific term, which is assumed normal with mean zero and variance σ_b^2 and independent for different i, c_t is a season-specific term, which is assumed normal with mean zero and variance σ_c^2 and independent for different t, and ε_{it} is the residual error term, assumed normal with mean zero and variance σ_ε^2 and independent for all data. After defining the variables $\frac{I_{jit}}{R_{it}} w_{jkit}$, we fit all of the β and σ^2 parameters of this model using the package lme4 for R.

Appendix B.
Cubist model rules and parameters

Table B1 Tables containing the fitted Cubist model rules and multiple regression parameters. The intercept for each regression is given by μ. The first of the two rows of multiple regression model parameters for each model rule are those fitted by the Cubist software; the second are those refitted in the framework of a linear mixed model that accounts for the correlation structure of the data, with NA denoting soil constraint parameters removed from the original Cubist model due to their having a positive effect on yield. Subscripts A, B and C denote soil depth, S and T denote the spatial and temporal components of VPD. Multiple regression parameters relating to sand and clay are given with these variables defined as proportions, not percentages

Rule #	Rule conditions	Multiple regression model							
	(a) Sodicity								
	VPD_S > 8.4	μ	VPD_T	$sdcty_C$	$sand_C$				
1	&	1.85	-0.22	-0.09	-0.34				
	$sand_C$ > 23.2	1.96	-0.19	-0.10	-0.50				
	VPD_S <= 8.4								
	&	μ	VPD_T	VPD_S	$sdcty_B$	$clay_B$	$clay_C$	$sand_B$	$sand_C$
2	$sand_C$ > 23.2	2.76	-0.29	-0.07	-0.16	-0.43	0.43	-0.70	0.23
	&	3.30	-0.29	-0.09	-0.17	-0.85	0.35	-1.11	0.25
	VPD_T > -2.5								
	$sand_C$ <= 23.2	μ	VPD_T	VPD_S	$sdcty_B$	$sdcty_C$	$clay_A$	$clay_B$	$sand_C$
3	&	3.74	-0.25	-0.08	-0.25	-0.21	-1.17	0.86	-2.64
	VPD_T > -2.5	4.10	-0.21	-0.11	-0.30	-0.14	-0.77	0.43	-3.11
		μ	VPD_T	VPD_S	$sdcty_B$	$sdcty_C$	$sand_C$		
4	VPD_T <= -2.5	5.25	-0.11	-0.25	-0.32	0.12	-0.96		
		5.94	0.04	-0.30	-0.23	NA	-0.69		

Rule #	Rule conditions	Multiple regression model														
	(b) Acidity															
	VPD_S > 8.5	μ	VPD_T	VPD_S	$acdty_A$	$acdty_C$	$alkty_A$	$alkty_B$	$clay_A$	$clay_B$	$clay_C$	$sand_A$	$sand_B$	$sand_C$		
1	&	1.96	-0.22	0.02	0.08	-0.38	0.33	-0.16	0.49	-0.35	-0.83	0.55	-0.22	-1.03		
	$sand_C$ > 25.3	2.34	-0.19	0.00	NA	-0.35	NA	-0.17	0.39	-0.41	-0.63	0.15	-0.16	-0.84		
	VPD_S <= 8.5	μ	VPD_T	VPD_S	$acdty_A$	$acdty_B$	$acdty_C$	$alkty_B$	$alkty_C$	$clay_A$	$clay_B$	$clay_C$	$sand_A$	$sand_B$	$sand_C$	
2	&	3.46	-0.30	-0.09	-0.11	0.21	-0.28	-0.42	0.14	-1.11	0.79	-0.38	-1.19	0.73	-0.50	
	VPD_T > -2.5	3.95	-0.30	-0.11	-0.13	NA	-0.20	-0.39	NA	-1.36	0.88	-0.55	-1.42	0.77	-0.65	
	$sand_C$ <= 25.3															
	&	μ	VPD_T	VPD_S	$sand_C$											
3	VPD_S > 8.5	3.52	-0.23	-0.08	-2.94											
	&	4.10	-0.18	-0.13	-3.10											
	VPD_T > -2.5															
		μ	VPD_T	VPD_S	$alkty_B$	$alkty_C$	$sand_C$									
4	VPD_T <= -2.5	5.03	-0.11	-0.24	-0.36	0.42	-1.18									
		5.88	0.05	-0.30	-0.18	NA	-0.85									

Rule #	Rule conditions	Multiple regression model									
	(c) Salinity										
1	$sand_C$ > 53.1 & VPD_S > 8.2 & VPD_T > -2.7	μ	VPD_T	VPD_S	$slnty_C$	$clay_A$	$clay_B$	$clay_C$	$sand_A$	$sand_B$	$sand_C$
		3.04	-0.23	-0.01	-0.09	0.19	-1.15	-0.28	0.82	-1.39	-1.03
		2.32	-0.22	0.01	-0.09	0.04	-1.75	1.38	0.27	-1.64	0.31
2	VPD_S > 8.2 & $sand_C$ <= 53.1 & VPD_T > -2.7	μ	VPD_T	VPD_S	$slnty_C$	$clay_A$	$clay_C$	$sand_A$	$sand_B$	$sand_C$	
		2.71	-0.21	0.00	-0.20	0.56	-0.91	0.99	-1.02	-2.17	
		3.22	-0.16	-0.06	-0.11	1.19	-1.35	1.52	-0.88	-2.51	
3	VPD_S <= 8.2 & VPD_T > -2.7	μ	VPD_T	VPD_S	$slnty_C$	$clay_A$	$clay_B$	$clay_C$	$sand_C$		
		2.93	-0.33	-0.07	-0.15	-0.50	0.66	-0.40	-0.50		
		3.47	-0.31	-0.13	-0.17	-0.57	0.73	-0.56	-0.56		
4	VPD_T <= -2.7	μ	VPD_T	VPD_S	$clay_C$	$sand_A$	$sand_B$	$sand_C$			
		4.88	-0.02	-0.21	-0.04	1.45	-1.93	-0.06			
		4.54	-0.06	-0.24	0.25	0.74	-0.52	-0.58			

Rule #	Rule conditions	Multiple regression model						
(d)	**Compaction**							
	$VPD_S > 8.1$	μ	VPD_T	VPD_S	$cmpctn_B$	$clay_B$	$clay_C$	$sand_C$
1	&	3.46	-0.22	-0.08	1.00	-0.60	-0.57	-1.36
	$VPD_T > -2.7$	3.15	-0.19	-0.09	NA	-0.11	-0.18	-0.80
	$VPD_S <= 8.1$	μ	VPD_T	$cmpctn_B$	$cmpctn_C$	$clay_B$	$clay_C$	$sand_C$
2	&	2.74	-0.32	-0.80	-1.50	0.69	-0.73	-0.76
	$VPD_T > -2.7$	3.13	-0.28	-0.98	-1.11	0.47	-1.19	-1.18
		μ	VPD_T	VPD_S				
3	$VPD_T <= -2.7$	6.22	-0.65	-0.58				
		5.71	-0.35	-0.46				

![GRDC Grains Research & Development Corporation]

Appendix C.
Sensitivity analysis results

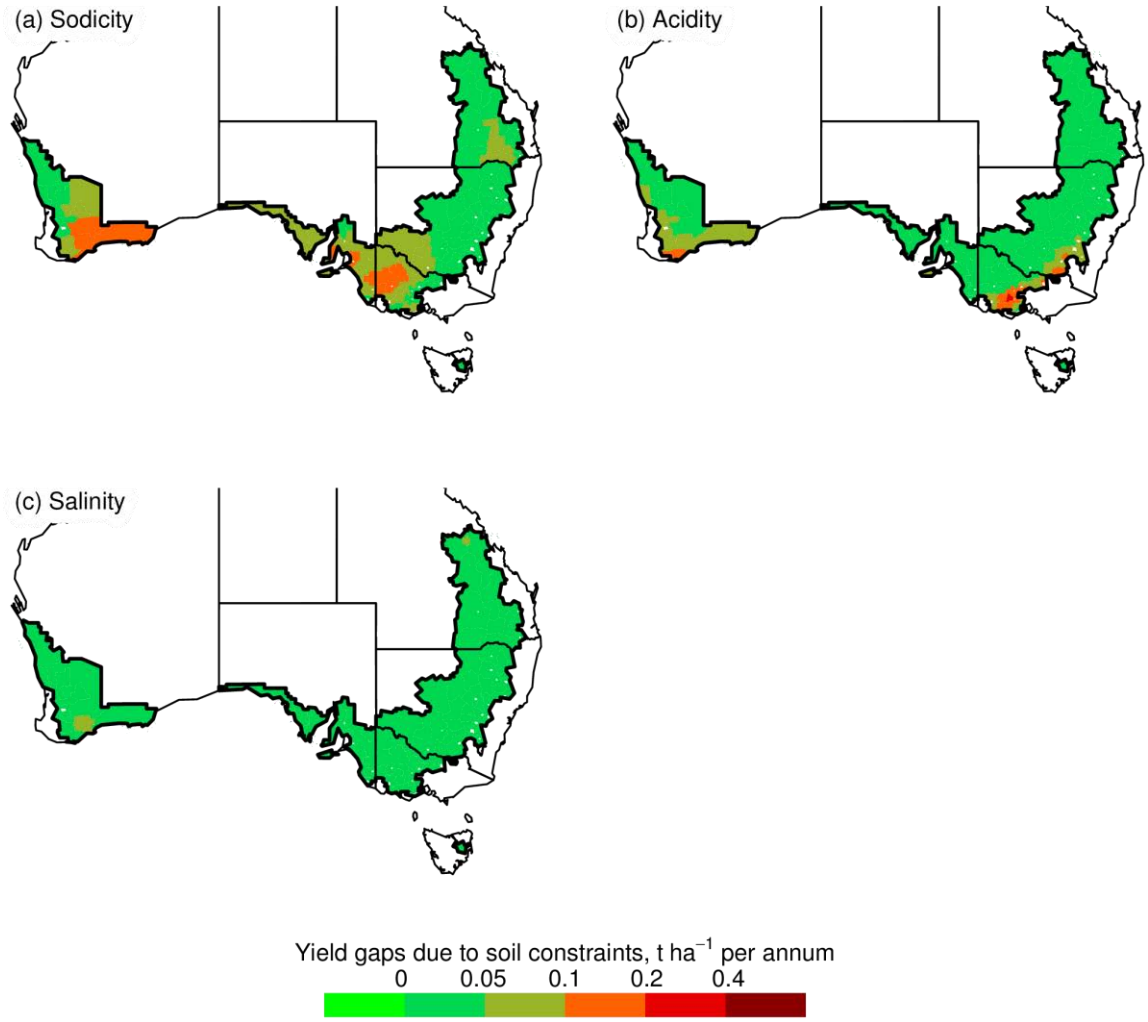


Figure C1 Yield gaps due to (a) sodicity, (b) acidity and (c) salinity at SA2 level based on the 'tolerant' critical values. Note that no map is presented for compaction since its tolerant critical values led to zero yield gaps

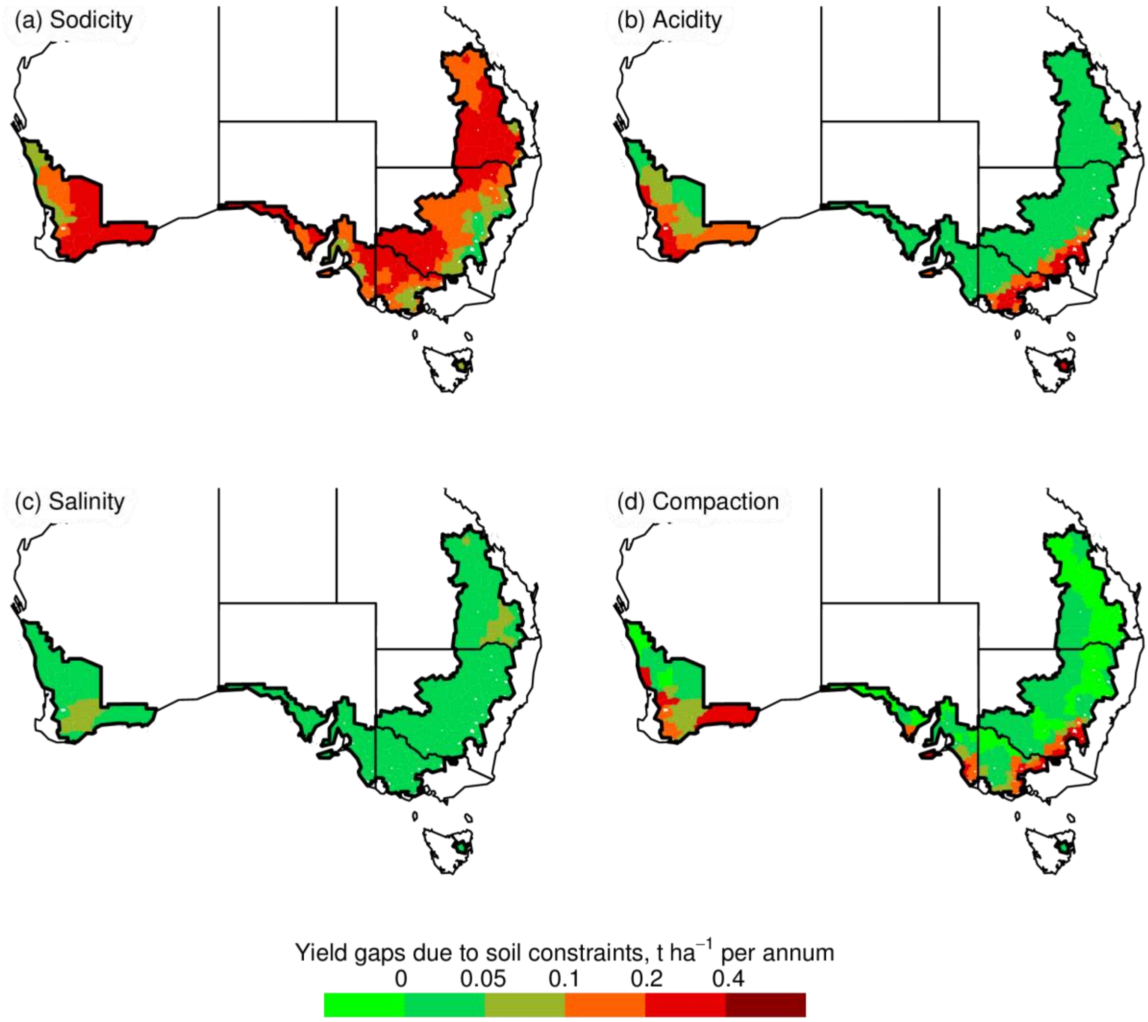


Figure C2 Yield gaps due to (a) sodicity, (b) acidity, (c) salinity and (d) compaction at SA2 level based on the 'sensitive' critical values

Table C1 Yield gap estimates with 'tolerant' critical values, summarized by state and nationally. pa: per annum

State	Area of cropping, kha	Soil constraint, *c*	Average yield gap due to *c*, t ha^{-1} pa	Area affected by *c*, kha	Gross value of yield gap due to *c*, M$ pa	Net value of ameliorating *c*, M$ pa
NSW	8064	Sodicity	0.03	3383	57.7	49.8
		Acidity	0.02	2267	37.7	32.6
		Salinity	0.01	754	19.2	NA
		Compaction	< 0.01	< 1	< 0.1	NA
Qld	3278	Sodicity	0.04	1873	34.2	29.5
		Acidity	0.01	183	4.3	3.7
		Salinity	0.03	876	22.1	NA
		Compaction	< 0.01	< 1	< 0.1	NA
SA	4964	Sodicity	0.08	3314	97.0	83.7
		Acidity	0.01	343	7.0	6.1
		Salinity	0.01	487	13.8	NA
		Compaction	< 0.01	< 1	< 0.1	NA
Tas	108	Sodicity	0.02	26	0.5	0.4
		Acidity	0.06	85	1.8	1.6
		Salinity	< 0.01	1	0.1	NA
		Compaction	< 0.01	< 1	< 0.1	NA
Vic	4614	Sodicity	0.10	3397	115.4	99.5
		Acidity	0.02	930	27.0	23.3
		Salinity	< 0.01	96	5.8	NA
		Compaction	< 0.01	< 1	< 0.1	NA
WA	17326	Sodicity	0.08	9325	352.7	304.2
		Acidity	0.04	9301	187.5	162.1
		Salinity	0.02	3262	81.8	NA
		Compaction	< 0.01	< 1	< 0.1	NA
		National summary				
	38354	Sodicity	0.07	21318	657.5	567.1
		Acidity	0.03	13109	265.3	229.3
		Salinity	0.01	5476	142.8	NA
		Compaction	< 0.01	< 1	< 0.1	NA

Table C2 Yield gap estimates with 'sensitive' critical values, summarized by state and nationally. pa: per annum

State	Area of cropping, kha	Soil constraint, *c*	Average yield gap due to *c*, t ha^{-1} pa	Area affected by *c*, kha	Gross value of yield gap due to *c*, M$ pa	Net value of ameliorating *c*, M$ pa
NSW	8064	Sodicity	0.15	6216	316.9	273.4
		Acidity	0.05	2048	97.0	83.8
		Salinity	0.01	1401	25.3	NA
		Compaction	0.03	1049	59.0	NA
Qld	3278	Sodicity	0.24	2861	202.4	174.6
		Acidity	0.02	584	15.8	13.7
		Salinity	0.04	1880	34.3	NA
		Compaction	< 0.01	37	0.9	NA
SA	4964	Sodicity	0.19	3903	243.3	209.9
		Acidity	0.01	752	18.1	15.7
		Salinity	0.01	971	18.6	NA
		Compaction	0.03	1134	32.8	NA
Tas	108	Sodicity	0.05	50	1.3	1.2
		Acidity	0.28	106	7.9	6.8
		Salinity	< 0.01	1	0.1	NA
		Compaction	0.02	47	0.5	NA
Vic	4614	Sodicity	0.23	3872	275.1	237.3
		Acidity	0.06	1388	71.4	61.7
		Salinity	0.01	191	7.4	NA
		Compaction	0.05	1062	54.4	NA
WA	17326	Sodicity	0.19	13948	834.6	719.9
		Acidity	0.11	10658	489.3	423.0
		Salinity	0.03	6609	154.9	NA
		Compaction	0.08	6262	350.7	NA
		National summary				
	38354	Sodicity	0.19	30850	1873.7	1616.1
		Acidity	0.07	15535	699.6	604.7
		Salinity	0.02	11053	240.5	NA
		Compaction	0.05	9591	498.2	NA

Appendix D.
APSIM modelling details

D1: Process overview

The following provides an overview of the process applied to set up APSIM so that it could reproduce (on average, and with the effects of soil constraints turned on) actual yields across Australia, and thereby simulate y_{ac} (with the NSSC data) and y_{oc} (with the constraint-optimized NSSC data).

1. First, characterize the general soil water holding capacities of seven soil orders using the data in APSoil (see Section D2 of this appendix); these representative soil water profiles were defined for 15-cm increments down to a depth of 30 cm, and for 30-cm increments from there down to 3 m.

Then, for each of the NSSC soil profiles with some ESP, pH, EC or Cl data:

2. Extract its soil order from the ASRIS map, and select the relevant soil water profile (of the seven representative profiles, see above)
3. Harmonize the depths of the NSSC profile to those of the selected soil water profile
4. Impose the effects of soil constraints on this profile: the effects of ESP, EC and Cl through APSIM's method of reducing the KL parameter according to a function described in Hochmann et al. (2007), and the effects of pH through a function that models its effect on the XF parameter (see Tang et al., 2003a and Tang et al., 2003b, and Section D3 of this appendix)
5. Extract the local climate data from the gridded climate database (SILO)
6. Run APSIM for this profile with five different wheat varieties (H45, Baxter, Yallaroi, Strzelecki and Sunbrook; a subset of those used in Hochmann et al., 2007) to compute the average yield over the years 1999–2012 for each variety

Once the above procedure has been carried out for all NSSC profiles and all five varieties, for each of the 250 wheat-growing SLAs:

7. Collate the simulated yields for all NSSC profiles falling in the SLA, and compute five average predicted yields, one for each of the five varieties. Select (as the representative variety for the SLA) the variety whose average predicted yield most closely matched the 1999-2012 average yield data for that SLA (data from ABS).

This gave us a set-up of APSIM that gave (on average) yields that were reasonably in line with actual yields. We then performed another set of APSIM simulations, with the selected variety for all NSSC profiles in a particular SLA but with each soil constraint in turn 'optimized' (i.e. any profiles with sodic soil were amended so that their ESP values were < 6 % through the profile). Equation (2) then gave us our yield gaps due to the soil constraints.

D2: Characterization of soil water capacity profiles for seven predominant soil orders

We used the APSoil data (http://www.apsim.info/) from 586 profiles to characterize the general soil water capacities of seven soil orders: Calcarosols, Chromosols, Dermosols, Kandosols, Sodosols, Tenosols and Vertosols. According to the ASRIS soil order and ACLUMP land use maps, these seven soil orders cover 95% of Australia's cropping land. Too few APSoil data were available representing the other seven Australian soil orders, so they were merged with their most similar soil order (from the seven most prevalent orders listed above) based on the minimum taxonomic distances between soil orders (Minasny and McBratney, 2007): Anthroposols, Organosols and Podosols were merged into the Dermosol soil order, Ferrosols were merged with the Chromosols, Hydrosols and Rudosols with the Tenosols, and the Kurosols with the Sodosols.

We classified all of the APSoil profiles according to the seven soil orders as according to the ASRIS soil order map, since it will be this map that determines to which of the seven orders each of the NSSC profiles belongs. For each of the soil orders in turn, all APSoil data classified as such were extracted and their crop lower limit for wheat (CLL) drained upper limit (DUL), saturation (SAT) and bulk densities (BD) modelled empirically using the linear mixed model approach described in Orton et al. (2016). This resulted in the seven general soil water characteristic curves are shown in Figure D1.

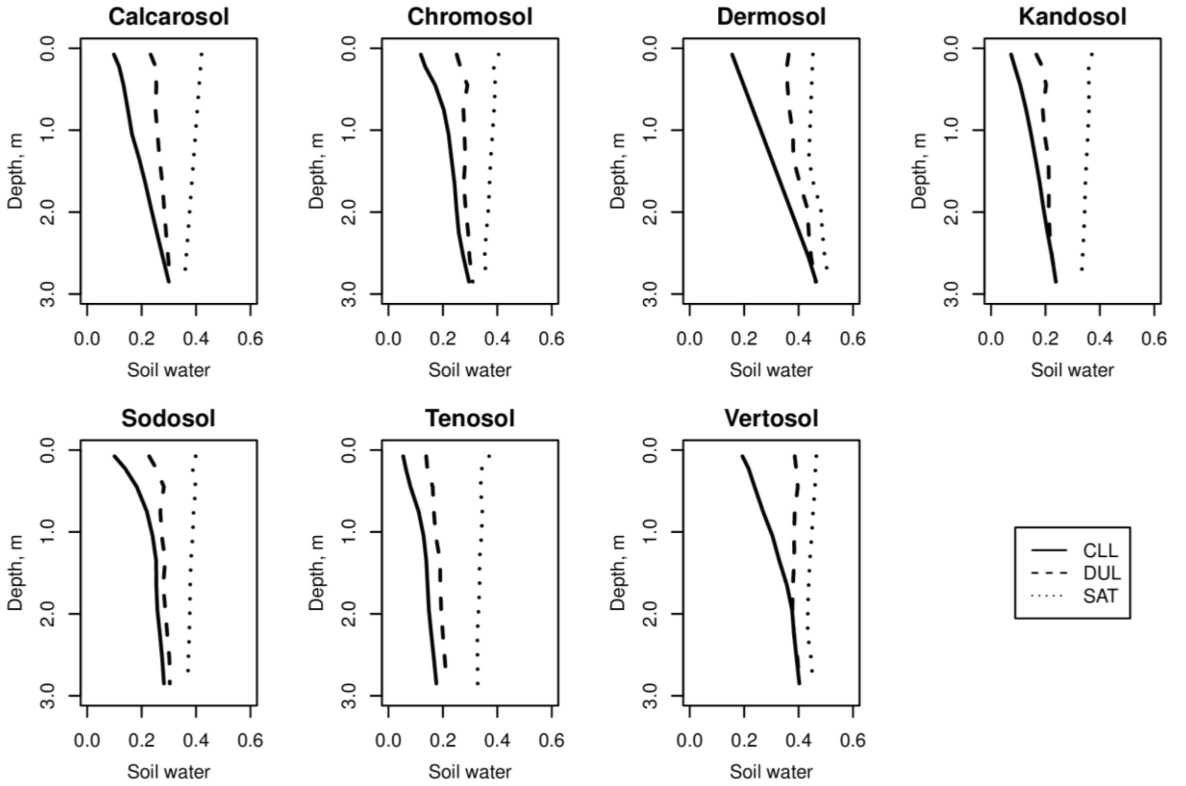


Figure D1 The general soil water curves (volumetric, and given as proportions) fitted to the Apsoil data for seven predominant soil orders covering 95% of Australia's wheat-cropping land: CLL: crop lower limit, DUL: drained upper limit, SAT: saturation

D3: Modelling the effects of soil acidity on the APSIM parameter XF

Tang et al. (2003b) investigated the effects of soil acidity on root length, and derived relationships between soil aluminium (Al) concentration and relative root length for Al-resistant and Al-sensitive wheat cultivars. In particular, the relationship between Al and relative root length (expressed as a proportion here, which effectively gives the XF parameter used by APSIM) for the Al-sensitive cultivar was given as:

$$XF = 0.0151 \exp\left(\frac{130}{31 + Al}\right)$$

Further, Tang et al. (2003a) presented data linking soil pH and Al concentration (extracted in 1:5 soil:0.01 M $CaCl_2$), which we extracted and modelled approximately as:

$$\ln Al = 17.6 - 3.65\ pH$$

To represent the effects of soil acidity, as measured by soil pH, in APSIM, we put these two relationships together to give a function relating soil pH to the APSIM parameter XF:

$$XF = 0.0151 \exp\left(\frac{130}{31 + \exp[17.6 - 3.65\ pH]}\right)$$

This function is shown in Figure D2.

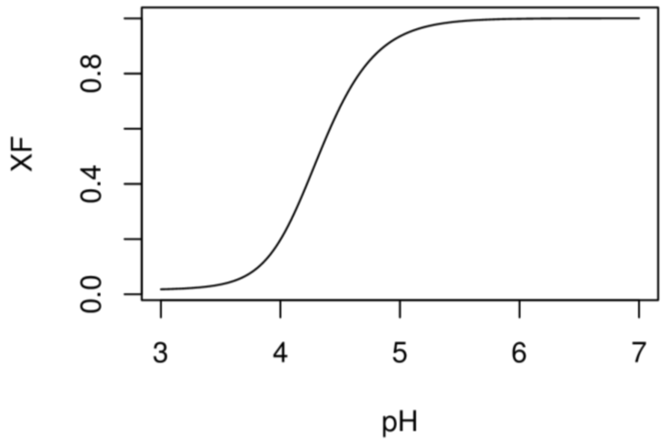


Figure D2 The function used to represent the effects of soil acidity, in which APSIM's root-exploration factor is modelled as a function of pH

Glossary and Acronyms

$A(r)$: the area of an SA2, indexed by *r*, used for cropping

ABS: Australian Bureau of Statistics

$acdty[d]$ or $acdty_d$: degree of acidity at depth *d*

ACLUMP: Australian Collaborative Land Use and Management Program

AIC: Akaike Information Criterion

$alkty[d]$ or $alkty_d$: degree of alkalinity at depth *d*

ASGS: Australian Statistical Geography Standard

ASRIS: Australian Soil Resource Information System

ATP kriging: Area-to-point kriging

BD: bulk density

$BDCrit_{Cmpctn}[d]$: critical value of BD for effects of compaction at depth *d*

c: used as an index for a soil constraint, representing sodicity, acidity, salinity or compaction

CEC: cation exchange capacity

Cl: chloride concentration

$ClCrit_{Slnty}[d]$: critical value of Cl for effects of salinity at depth *d*

$cmpctn[d]$ or $cmpctn_d$: degree of compaction at depth *d*

EC: electrical conductivity

$ECCrit_{Slnty}[d]$: critical value of EC for effects of salinity at depth *d*

ESP: exchangeable sodium percentage

$ESPCrit_{Sdcty}[d]$: critical value of ESP for effects of sodicity at depth *d*

EVI: enhanced vegetation index

GRDC: Grains Research and Development Corporation

iEVI: time-integrated enhanced vegetation index

LMM: linear mixed model

lnESP: the log of the ESP

lnCl: the log of the chloride concentration

NDVI: normalized difference vegetation index

NSSC: National Soil Site Collation

$P_{gc}(r)$: average probability of land being affected by soil constraint *c* in an SA2 indexed by *r*

$P_{gc}(\mathbf{x}_\mathrm{k})$: indicator variable, indicating if yield gap due to soil constraint *c* is greater than zero at the interpolation grid locations, $\mathbf{x}_\mathrm{k}$

$P_{gc}(\mathbf{x}_\mathrm{S})$: indicator variable, indicating if yield gap due to soil constraint *c* is greater than zero at the soil data locations, $\mathbf{x}_\mathrm{S}$

pH_{Ca}: pH measured in a 1:5 0.01 M calcium chloride solution

$pHCrit_{Acdty}[d]$: critical value of pH for effects of acidity at depth *d*

$pHCrit_{Alkty}[d]$: critical value of pH for effects of alkalinity at depth *d*

pH_{W}: pH measured in a 1:5 soil:water solution

REML: restricted maximum likelihood

SA2: Statistical Area Level 2

SAR: sodium adsorption ratio

$sdcty[d]$ or $sdcty_d$: degree of sodicity at depth *d*

SILO: Scientific Information for Land Owners

SLA: Statistical Local Area

$slnty[d]$ or $slnty_d$: degree of salinity at depth *d*

Soil depth A: 0–10 cm

Soil depth B: 10–50 cm

Soil depth C: 50–200 cm

VPD: vapour pressure deficit

VPD_S: spatial (long-term average) component of VPD

VPD_T: temporal (yearly departure from the long-term average) component of VPD

Y_a: actual yield, as used in Hochmann et al. (2012)

Y_{ac}: actual yield, as predicted by model with actual soil data

Y_g: yield gap, as defined in Hochmann et al. (2012)

Y_{gc}: yield gap due to soil constraint *c*

$Y_{gc}(r)$: average yield gap due to soil constraint *c* in an SA2 indexed by *r*

$Y_{gc}(\mathbf{x}_\mathrm{k})$: yield gap due to soil constraint *c* at the interpolation grid locations, $\mathbf{x}_\mathrm{k}$

$Y_{gc}(\mathbf{x}_\mathrm{S})$: yield gap due to soil constraint *c* at the soil data locations, $\mathbf{x}_S$

Y_{oc}: constraint-optimized yield, as predicted by model with soil constraint *c* optimized

Y_w: water-limited potential yield, as defined in Hochmann et al. (2012)

References

AACM International, 1995. Social and economic feasibility of ameliorating soil acidification. A national review for the Land and Water Resources Research and Development Corporation.

ABS, 2017. Australian Statistical Geography Standard (ASGS). http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+(ASGS)

Akaike H, 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.

Allen DE, Pringle MJ, Page KL, Dalal RC, 2010. A review of sampling designs for the measurement of soil organic carbon in Australian grazing lands. The Rangeland Journal 32, 227–246.

Ashton D, Oliver M, Valle H, 2016. Australian grains: financial performance of grain farms, 2013– 14 to 2015– 16. Canberra: ABARES.

Bunting P, Clewley D, Lucas RM, Gillingham S, 2014. The Remote Sensing and GIS Software Library (RSGISLib). Computers and Geosciences 62, 216–226.

Bot AJ, Nachtergaele F, Young A, 2000. Land resource potential and constraints at regional and country levels. World Soil Resources Reports 90, 114.

Dang YP, Dalal RC, Buck SR, Harms B, Kelly R, Hochman Z, Schwenke GD, Biggs AJW, Ferguson N, Norrish S, Routley R, McDonald M, Hall C, Singh J, Daniells IJ, Farquharson R, Manning W, Speirs S, Grewal HS, Cornish P, Bodapati N, Orange D, 2010. Diagnosis, extent, impacts and management of subsoil constraints in the northern grains cropping region of Australia. Australian Journal of Soil Research 48, 105–119.

Dang YP, Moody PW, 2016. Quantifying the costs of soil constraints to Australian agriculture: a case study of wheat in north-eastern Australia. Soil Research 54, 700–707.

DERM, 2011. Salinity management handbook, second edition. Department of Environment and Resource Management, Queensland. https://publications.qld.gov.au/storage/f/2013-12-19T04%3A10%3A23.754Z/salinity-management-handbook.pdf

DNRM, 2007. Constraints to cropping soils in the northern grains region: A decision tree. Queensland Department of Natural Resources & Water. http://www.grdc.com.au/uploads/documents/dnr00004.pdf

Don A, Schumacher J, Scherer-Lorenzen M, Scholten T, Schulze ED, 2007. Spatial and vertical variation of soil carbon at two grassland sites – implications for measuring soil carbon stocks. Geoderma 141, 272–282.

Flood N, Danaher T, Gill T, Gillingham S, 2013. An Operational Scheme for Deriving Standardised Surface Reflectance from Landsat TM/ETM plus and SPOT HRG Imagery for Eastern Australia. Remote Sensing 5, 83–109.

Gazey C, Davies S, 2009. Soil acidity: A guide for WA farmers and consultants. The department of Agriculture and Food, Western Australia.

Hajkowicz S, Young M, 2005. Costing yield loss from acidity, sodicity and dryland salinity to Australian Agriculture. Land Degradation and Development 16, 417–433.

Henderson BL, Bui EN, 2002. An improved calibration curve between soil pH measured in water and $CaCl_2$. Australian Journal of Soil Research 40, 1399–1405.

Hochman Z, Dang YP, Schwenke GD, Dalgliesh NP, Routley R, McDonald M, Daniells IG, Manning W, Poulton PL, 2007. Simulating the effects of saline and sodic subsoils on wheat crops growing on Vertosols. Australian Journal of Agricultural Research 58, 802–810.

Hochman Z, Gobbett D, Holzworth D, McClelland T, van Rees H, Marinoni O, Garcia JN, Horan H, 2012. Quantifying yield gaps in rainfed cropping systems: A case study of wheat in Australia. Field Crops Research 136, 85–96.

Huete AR, Liu HQ, Batchily K, van Leeuwen W, 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. Remote Sensing of Environment 59, 440–451.

Hughes N, Lawson K, Davidson A, Jackson T, Sheng Y, 2011. Productivity pathways: climate adjusted production frontiers for the Australian broadacre cropping industry. ABARES research report 11.5.

Jeffrey SJ, Carter JO, Moodie KB, Beswick AR, 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environmental Modelling and Software 16, 309–330.

Kokic P, Davidson A, Rodriguez V, 2006. Australia's grains industry: factors influencing productivity growth. Australian Commodities 13, 705–712.

Kyriakidis PC, 2004. A Geostatistical Framework for Area-to-Point Spatial Interpolation. Geographical Analysis 36, 259–289.

Lobell DB, 2013. The use of satellite data for crop yield gap analysis. Field Crops Research 143, 56–64.

Mallawaarachchi T, Walker P, Young MD, Smyth RE, Lynch HS, Dudgeon G, 1996. GIS-based integrated modeling systems for natural resource management. Agricultural Systems 50, 169–189.

McKenzie N, Jacquier D, Isbell R, Brown K, 2004. Australian Soils and Landscapes: An illustrated compendium. CSIRO Publishing, Australia.

Minasny B, McBratney AB, 2007. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. Geoderma 142, 285–293.

Minasny B, McBratney AB, Brough DM, Jacquier D, 2011. Models relation soil pH measurements in water and calcium chloride that incorporate electrolyte concentration. European Journal of Soil Science 62, 728–732.

NLWRA (National Land and Water Resources Audit), 2001. Australian agriculture assessment. NLWRA: Canberra, ACT.

Northcote KH, Skene JKM, 1972. Australian soils with saline and sodic properties. CSIRO Soil Publication No. 27, CSIRO Melbourne, Australia.

Odeh IOA, McBratney AB, Chittleborough DJ, 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67, 215–226.

Orton TG, Pringle MJ, Bishop TFA, 2016. A one-step approach for modelling and mapping soil properties based on profile data sampled over varying depth intervals. Geoderma 262, 174–186.

Orton TG, Román Dobarco M, Saby NPA, 2017. Kriging based on areal summary statistics data: Effects of within-unit variability on predictions and uncertainties. Spatial Statistics 19, 42–67.

Patterson HD, Thompson R, 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545–554.

Petersen E, 2015. Economic analysis of the impacts and management of subsoil constraints. Perth: Department of Agriculture and Food, Western Australia.

Pinheiro JC, Bates DM, 2004. Mixed-effects models in S and S-PLUS. Springer Science & Business Media, New York, USA.

Pringle MJ, Allen DE, Orton TG, Bishop TFA, Butler DW, Henry BK, Dalal RC, 2016. Effects of land-use change and management on soil carbon and nitrogen in the Brigalow Belt, Australia: II. Statistical models to unravel the climate-soil-management interaction. The Rangeland Journal 38, 453–466.

Quinlan R, 1992. Learning with continuous classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence 92, 343–348.

Rayment GE, Lyons DJ, 2011. Soil chemical methods–Australasia. CSIRO Publishing, Australia.

Rural Solutions SA, 2016. Farm Gross Margin Guide: A gross margin template for crop and livestock enterprises. Adelaide: A Rural Solutions SA.

Searle R, 2014. The Australian site data collation to support the GlobalSoilMap. In GlobalSoilMap: Basis of the global spatial soil information system, Eds: Arrouays D, Mckenzie N, Hempel J, de Forges ACR, McBratney A. CRC Press, The Netherlands.

Sheng Y, Jackson T, Gooday P, 2017. Resource reallocation and its contribution to productivity growth in Australian broadacre agriculture. Australian Journal of Agricultural and Resource Economics 61, 56–75.

Soil Science Australia, 2015. Measuring soil cation exchange capacity and exchangeable cations. http://www.soilscienceaustralia.com.au/images/sampledata/branches_tab/queensland/occasionalpublications/measuring_soil_cation_exchange_capacity_and_exchangeable_cations.pdf

Tang C, Rengel Z, Diatloff E, Gazey C, 2003a. Responses of wheat and barley to liming on a sandy soil with subsoil acidity. Field Crops Research 80, 235–244.

Tang C, Asseng S, Diatloff E, Rengel Z, 2003b. Modelling yield losses of aluminium-resistant and aluminium-sensitive wheat due to subsurface soil acidity: effects of rainfall, liming and nitrogen application. Plant and Soil 254, 349–360.

Tetens O, 1930. Uber einige meteorologische Begriffe. z. Geophys. 6:297-309.

Upjohn B, Fenton G, Conyers M, 2005. Soil acidity and liming. Ed. NSW Department of Primary Industries. http://www.dpi.nsw.gov.au/

Van Gool D, 2016. Identifying soil constraints that limit wheat yield in the south-west of Western Australia. Department of Agriculture and Food, Western Australia, Perth. Report RMTR 399.

Webster R, Oliver MA, 2007. Geostatistics for Environmental Scientists. John Wiley & Sons, Chichester, UK.

Werker AR, Jaggard KW, 1997. Modelling Asymmetrical Growth Curves that Rise and then Fall: Applications to Foliage Dynamics of Sugar Beet (*Beta vulgaris* L.). Annals of Botany 79, 657–665.

Social Media Posting

GRDC uses social media to showcase research investments and disseminate timely, relevant and practical information to key stakeholders in the grains industry. Our audiences are predominantly growers and agricultural advisers.

Social Media Accounts:

Facebook: https://www.facebook.com/theGRDC
Twitter: https://twitter.com/theGRDC
YouTube: http://www.youtube.com/user/theGRDC
LinkedIn: http://www.linkedin.com/company/thegrdc

Is there any reason why this report cannot be communicated on social media?

a. *GRDC have advised that it will be to their discretion whether this report is communicated on social media.*

If no, please provide the following:

1. Who is the target audience for this content? (e.g., growers, adviser, researchers, policy makers, etc.)

    a. *GRDC*

2. At what time of year is this content most relevant to the target audience?

    a. *Any*

3. On which of GRDC's social media accounts would you like this content posted? Please provide text (2-3 sentences for Facebook and LinkedIn and 140 characters for Twitter), images, graphs, or charts that support the content. Where applicable, please include any relevant Twitter handles (usernames) for project staff.

    a. To the discretion of GRDC